

Promotion Letters: Current Problems and a Reform Proposal

Kurt Weyland, *University of Texas at Austin*

ABSTRACT External referee letters can, in principle, provide essential inputs for tenure and promotion decisions. But scholars' reluctance to prepare critical, not to speak of "negative" assessments has led to a skewed self-selection: Potential referees tend to accept the task only if they see a tenure or promotion case as worthy of support. The resulting overabundance of praise has devalued the content of referee reports; they are difficult to interpret and of limited use for distinguishing deserving from undeserving candidates. Maybe a simple reform can mitigate these problems: Departments should pay external referees a generous honorarium, perhaps up to \$2,000 at well-endowed research universities. This substantial incentive will induce more scholars to provide these assessments, thus counterbalancing the current self-selection on candidate supportability. Moreover, a large honorarium clearly turns the commissioning department into the contractual "principal" of these assessments, leading referees to prepare more searching, balanced evaluations. The main criterion will be whether a promotion really improves the department's academic quality and scholarly standing.

As an essential input for tenure and promotion decisions, universities commission external-referee letters that are supposed to assess the quality and heft of a candidate's contribution to their field of scholarship. Certainly, a thorough analysis of the CV should be the fundamental basis for making these high-stakes decisions. Moreover, citation counts assess the impact of a candidate's publications. However, these objective measures are not perfectly valid indicators of academic quality and scholarly promise, especially for assistant professors with a limited track record.

Therefore, universities complement the CV and citation count with assessments by outside scholars. Academic experts can best judge the substance and quality of a candidate's contribution to scholarship. They can evaluate how much the published work has advanced knowledge and understanding; how innovative and creative this research has been; and how promising the projects for the future are. Is there a coherent research program with a logical progression from more specific investigations toward broader studies?

REFeree SELF-SELECTION AND ITS PROBLEMS

Recent tendencies, however, make it ever more doubtful that external-referee reports properly fulfill these important functions. Arguably, there has been a "devaluation" of outside evaluations in their content and contribution to academic decision making.

Kurt Weyland is professor of government at the University of Texas at Austin. He may be reached at kweyland@austin.utexas.edu.

Because professional confidentiality precludes precise measurement of this erosion, this article cannot draw on systematic data analysis. However, conversations with various colleagues and long-standing participation in departmental decision making, service on my college's promotion and tenure committee, and the preparation of frequent external-referee reports provide significant informal, qualitative evidence. Comments received during the double-blind review process corroborate these impressions.

Referee letters have diminished in academic value because their tenor has become overwhelmingly positive. Explicit criticism is infrequent and negative recommendations have become rare. Most assessments are full of praise, and recommendations on tenure and promotion are almost uniformly supportive.

Arguably, this positive skew is not a faithful reflection of the actual strength of promotion cases. Rational self-selection by candidates is not perfect enough to guarantee that only strong, deserving assistant and associate professors undergo the process. Whereas hopeless candidates tend to withdraw, the high rewards of tenure induce many assistant professors who lack clear-cut, unchallengeable records to try; and associate professors well may be tempted to bid for promotion before they have established a strong record. Departments and universities, in fact, must face quite a number of borderline cases. Consequently, one would expect a good share of mixed referee reports and a number of negative verdicts. Compared to this reasonable prediction, however, the promotion letters that departments actually receive have a strikingly positive slant.

This lopsided balance sheet suggests that many senior scholars are reluctant to offer frank professional assessments that truly reflect the strengths and weaknesses of tenure and promotion cases. There seems to be increasing self-selection among potential referees: Scholars tend to accept this task only if they perceive a candidate as worthy of support. When approached about a weak or questionable case, senior scholars invoke prior commitments or lack of time (I have done this myself!) and deliberately avoid the problematic job of judging an undeserving or borderline case. Due to this self-selection, the assessments that referees do write are predominantly positive and thorough criticism remains rare.

There are powerful reasons for the hesitation of established academics to judge potentially problematic cases and to be properly critical in tenure and promotion evaluations. One concern may arise from the open-records laws that apply especially to public universities. In fact, in this era of litigation, it may not be clear to potential referees whether aggrieved candidates would be denied access to any of their written assessments. Given the trouble that a legal challenge might cause, senior scholars may cautiously avoid critical comments and confine themselves to praise.

The Zeitgeist pervading contemporary US academia poses additional obstacles to honest, potentially critical judgments, especially to explicit recommendations to deny tenure and promotion.

The *Zeitgeist* pervading contemporary US academia poses additional obstacles to honest, potentially critical judgments, especially to explicit recommendations to deny tenure and promotion. In this age of anxiety, scholars seem ever more reluctant to criticize colleagues directly, in writing, and with their signature. This hesitation is especially pronounced because—by contrast to peer reviews of academic works, which often are quite critical—a tenure and promotion assessment focuses directly on a specific person and can seriously affect their life, considering the recent tight job market. Why undertake the major effort to elaborate a thorough, comprehensive evaluation of a weak candidate and arrive at a negative recommendation, especially in a tenure case? Who wants to cost a colleague their job, especially somebody one knows personally? An informal norm seems to be emerging that declares a critical assessment—especially a negative tenure and promotion assessment—as inappropriate and illegitimate. Given that the potential referee can easily arrive at a “ballpark” assessment via a candidate’s CV, the *Zeitgeist* leads senior scholars to decline evaluations of weak candidates. Instead, they tend to review only those cases that they can support.

The academic self-interest of established scholars—that is, their own quest for new knowledge and understanding—intensifies these tendencies toward self-selection. After all, a strong tenure and promotion case is more interesting and less time-consuming to evaluate than a weak case. Although meritorious candidates usually have larger publication records than weak candidates, their major contributions—which are sufficient for justifying a positive recommendation—are not difficult to identify. A referee can explain and highlight these contributions by focusing on the main publications, which are also instructive and beneficial to read; in fact, an expert often will already know these works. What a pleasure to engage the work of a young hotshot! To evaluate a

weak candidate, by contrast, the conscientious referee would feel compelled to digest virtually all the extant writing, which usually is not very good. After all, before drafting a mixed evaluation and especially before arriving at a negative verdict, the referee must be sure not to overlook a nugget buried somewhere in these publications, manuscripts, and drafts. For this purpose, an expert would have to read a good deal of prose that most likely is not very enlightening. Why spend so much time on an unpleasant task? For their own scholarly benefit, leading experts will be tempted to decline evaluating problematic candidates.

THE DEVALUATION OF REFEREE REPORTS

Due to this self-selection based on candidate supportability, a large majority of external-referee reports is positive. Praise clearly predominates, whereas reservations remain limited and criticisms usually are presented as afterthoughts, tucked away in dependent clauses. Above all, virtually every outside assessment ends with a positive verdict on tenure and promotion. Even most of the few evaluations that present a more balanced set of pros and cons conclude with an unbalanced bottom line—namely, a thumbs-up.

Certainly, this self-selection based on supportability might not diminish the quality of tenure and promotion decisions if all external referees applied equally high professional standards. In this case, departments would manage to commission the requisite number of evaluations only for supportable candidates. Weak cases clearly would fail because experts would refrain from providing a referee report. Consequently, numerous tenure and promotion processes would be terminated early.

However, that is a rare occurrence. Few tenure and promotion cases fail because a department cannot obtain a sufficient number of external-referee reports. Evidently, established scholars differ quite starkly in their professional standards: Some are committed to stringent criteria, others less so, and some colleagues are generous and lenient. Academia is diverse and heterogeneous; there are clear professional divergences concerning the standards that our discipline should apply. Some of these disagreements arise from differences in theoretical commitments and methodological persuasions, which are common in our pluralistic field. Moreover, there are differences in preferred academic style: Some colleagues value frequent contributions to normal science, whereas others believe that a seminal breakthrough can make a career.

Yet, there also are substantial differences in the stringency of the professional criteria that political scientists apply; some colleagues are clearly less demanding than others. Due to these different levels of professional standards, departments manage to find enough outside reviewers even for many weak candidates. Whereas the widely recognized leaders in a candidate’s broader field may not want to write, scholars in the specific area of specialization may be willing to do so—or at least colleagues in a candidate’s narrow niche.

Given these divergent standards, the previously mentioned self-selection ensures that even referee reports about problematic

or weak promotion cases tend to be overwhelmingly positive. If candidates have not published much, then their manuscripts can be praised. After all, how are terms such as “excellent” defined, exactly? As with grade inflation in many colleges, there has been an increasingly liberal use of categories of distinction. Of course, the proliferation of relative praise undermines its meaning. If so many scholars are outstanding, then no one really stands out. Due to this widespread debasement of the currency, the content of external-referee reports is difficult to read. These assessments have suffered a considerable devaluation as input into high-stakes, up-or-down decisions.

If so many scholars are outstanding, then no one really stands out. Due to this widespread debasement of the currency, the content of external-referee reports is difficult to read. These assessments have suffered a considerable devaluation as inputs into high-stakes, up-or-down decisions.

In addition to this combination of self-selection and divergent professional standards, there can be special reasons for a referee to support a borderline candidate. There are rumors of letter-writing networks through which (groups of) senior scholars try to ensure that their former students advance in their academic careers. Also, given the methodological and theoretical divisions in our discipline, potential referees may be concerned that a borderline candidate might not get a fair hearing in their home department. Perhaps a colleague does the type of research or publishes in the kind of outlets that the home department does not appreciate, yet the outside expert has a more sanguine view of the candidate’s work and these publication venues. In this situation, a senior scholar may accept the evaluation task and, to counterbalance the perceived bias of the candidate’s home department, may highlight the strengths and downplay the weaknesses of the case. Of course, particularly strong theoretical or methodological commitments or other bases of solidarity may induce outside experts to overestimate the merit of a candidate whom they would like to support. These motivations for self-selection benefit weaker candidates, but they contribute to the positive bias of extant referee reports and their tendency to deviate from frank, honest evaluations.

As a result of these mixed motives, the tenure and promotion assessments that departments receive from outside experts are difficult to interpret. Does their overwhelmingly positive tenor reflect the true quality of a specific candidate’s work? Are the few critical comments that do appear merely tactical efforts to signal balance and claim credibility—or are they the “tip of the iceberg,” faintly indicating major reservations that the external referee mostly chooses to suppress? “Reading between the lines” is a difficult task that does not reliably yield valid results. After all, what should count more: the “waffle words” and ambiguous phrases sprinkled into the main body of an assessment, or the positive vote that almost invariably appears at the end? The only thing that is clear is that the content of many external-referee reports cannot be taken at face value.

Consequently, departmental and university decision makers shift their attention to other aspects of the external-referee process. What matters more and more is the number of experts

a department must approach in order to obtain a sufficient number of reports. It is not so difficult to persuade leading scholars to evaluate the work of a strong candidate; but how difficult it can be to round up enough assessments of problematic or weak candidates! In some cases, this acceptance rate falls well below 50%. Of course, what also tend to decline in weak cases are the caliber of the external referees and the standing of their institutions. Academic “stars” are willing to write about excellent candidates but have no interest in evaluating problematic files; for the latter, departments often succeed in commissioning only scholars of modest prestige from a candidate’s niche and circle of close

colleagues. Thus, because the tendency toward uniform praise has devalued the content of external-referee reports, departments’ inferences about the quality of a tenure and promotion candidate are shifting to the acceptance rate for outside evaluations, the prominence of the scholars who do write, and the rank of their institutions.

These indirect indicators are suboptimal, however, compared to the role that thorough, searching evaluations of the strengths and weaknesses of a candidate’s research and publications could play. The caliber of referees and their university’s prestige constitute imperfect proxies that provide much less detailed and relevant information. Moreover, whereas the acceptance rate for referee reports clearly deserves consideration, it provides uncertain evidence. Some outside scholars do have compelling reasons to decline. Furthermore, because some experts commit to self-restraint and accept only a limited number of these assignments per year, chance factors can distort acceptance rates. A department that is late in commissioning referee reports can unfairly saddle a deserving candidate with a low acceptance rate. Because scholars at top institutions are especially sought out and quickly “booked up,” such a delay can artificially produce an unimpressive list of evaluators who eventually do write. In summary, given that potential referees have variegated reasons for declining the task, indirect inferences via acceptance rates and referee prestige form an uncertain and problematic base for decisions that are of crucial importance, for both the candidates and their universities.

A REFORM PROPOSAL

This article argues that external-referee reports have been devalued by a predominance of praise. For departments, it certainly would be preferable if these assessments reflected more honestly the strengths and weaknesses of tenure and promotion candidates. For this purpose, it would be crucial to weaken or counteract the tendency toward self-selection on candidate supportability that has skewed extant outside evaluations in such a pronounced fashion.

How can this problem be alleviated? It is not easy to think of promising, feasible solutions. Appeals to professional norms are unlikely to effect significant and lasting change. APSA training

courses may not elicit much interest, and a system of reviewer certification would be difficult to establish across our discipline's different subfields.

For purposes of discussion, therefore, I propose a simple reform: Departments should offer outside experts a generous honorarium (e.g., perhaps up to \$2,000 at leading, well-endowed research

provide these assessments and to base them on the priorities of the commissioning university; they would no longer accept the task only when they can support the candidate. These assessments thus would become a contractual assignment rather than a favor for the candidate. As a result, external referee letters would provide better, more professional inputs for high-stakes decisions

For purposes of discussion, therefore, I propose a simple reform: Departments should offer outside experts a generous honorarium (e.g., perhaps up to \$2,000 at leading, well-endowed research universities) for a thorough, searching assessment, especially a tenure review.

universities) for a thorough, searching assessment, especially a tenure review. This substantial monetary incentive would have two main purposes. First, it would be designed to counterbalance the currently prevailing self-selection among potential evaluators. The goal is to induce a broader set of scholars—including those who are not necessarily predisposed to support a candidate—to accept the time-consuming task of elaborating an assessment. A major financial reward likely would extend the range of referees and lead scholars with higher academic standards to accept a wider range of assignments, even in borderline and perhaps negative cases. For instance, a large honorarium would compensate for the greater time and effort it takes to evaluate a weak tenure and promotion case. Money talks—and that is a good thing in this case! A substantial payment can motivate referees even if a candidate does not have an unchallengeable publication record or if the referee does not have other reasons to support the candidate. Thus, a generous honorarium would balance out the range of evaluators and extend it beyond the set of self-selected supporters. Consequently, there is a good chance that departments would receive more even-handed, honest assessments.

Second, a generous honorarium would probably induce this broader range of referees to elaborate more thorough, searching, and honest reports. The main motivation for accepting the task would shift beyond support for the candidate. Because the commissioning department would clearly become the principal, these evaluations would become contractual assignments that would probably produce more objective judgments of candidate quality. Through the high reward, this department would commit more referees to taking its main goals and interests as the fundamental yardstick of the assessment: Will this department really benefit significantly from a positive tenure and promotion decision? The spirit of external evaluations likely would shift from supporting the candidate toward a more neutral, objective weighing of the candidate's strengths and weaknesses. Through a large honorarium, the candidate's university would take full ownership and gain the undeniable right to receive thorough, professional evaluations, which would fairly measure accomplishments and promise—or the lack thereof. External experts would incur the contractual obligation to orient these assessments toward the commissioning university's needs and priorities—especially the crucial institutional interest of distinguishing between deserving and undeserving candidates—and to disregard the factors that currently favor the candidate.

In summary, a generous honorarium—with the specific level depending on an institution's financial capacity and research orientation—gives a wider range of scholars good reasons to

than current reports do, which inflationary praise has devalued. Thus, an injection of money could revalue a debased currency and yield a more worthwhile product for the commissioning department.

From a university's perspective, the expense arising from this proposal represents a prudent investment that is professionally justified. After all, a tenure decision involves a financial commitment of more than one million dollars; therefore, it makes sense for a well-endowed institution to spend \$12,000 to \$16,000—about one percent of the prize to be awarded—on obtaining high-quality expert assessments that help improve these decisions. The hiring process for an assistant professor, which typically involves interviewing three job candidates, costs approximately \$5,000. Given that an assistant professorship lasts six years, whereas the average academic career extends to about 35 years, universities are well advised to spend the additional sum on deciding whether to grant an assistant professor permanent employment. For these reasons, administrators with a long-term perspective and a commitment to academic excellence may find this proposal of a generous honorarium cost-effective.

Of course, this simple reform plan will not solve all problems. Moreover, it has some downsides, such as exacerbating inequality among universities. However, the higher honorarium that top institutions would offer is justified by the larger average publication record of their promotion candidates and, especially, the greater stakes involving these plum jobs. Absent feasible alternatives, the honorarium proposal may be the best step forward.

CONCLUSION

This article argues that external-expert assessments have lost much of their value for tenure and promotion decisions. Pronounced self-selection among potential referees has led to a proliferation of praise and an avoidance of critical comments. Because scholars who do write these reports use differential standards and because a number of referees have specific reasons to support candidates, this predominance of praise does not faithfully reflect the accomplishments and scholarly standing of assistant and associate professors. Therefore, the content of these evaluations has become problematic as an indicator of academic quality.

To compensate for this problem, decision makers have begun to rely more on other indicators, especially the caliber of the scholars who are willing to write and how many experts must be contacted to guarantee a sufficient number of reports. However, these alternatives also are plagued by deficiencies; although they offer a “ballpark” idea about a candidate's standing, they provide

a much less in-depth analysis of the scholarly work and the candidate's academic promise for the future.

In light of these difficulties, this article proposes that departments offer generous honoraria for external-expert assessments. By shifting the base of self-selection away from the perceived supportability of a candidate, a large monetary incentive can broaden the range of reviewers, thereby leading to more searching, honest evaluations. Moreover, the academic interests and needs of the commissioning department would be more likely to orient and guide these evaluations. Given the enormous stakes of promotion

decisions, including the financial commitments involved in tenure, universities are well advised not to penny-pinch—and to improve their decision making in this way.

ACKNOWLEDGMENTS

I thank Wendy Hunter, Bartholomew Sparrow, and three anonymous reviewers for important comments on previous versions. This article was presented at APSA's 112th Annual Meeting in Philadelphia in September 2016. I am grateful to Bartholomew Sparrow for organizing this panel and the current spotlight. ■