Available online at www.sciencedirect.com
SCIENCE DIRECT®
Poetics 31 (2003) 23–49
POETICS
www.elsevier.com/locate/poetic

From character to intellect: changing conceptions of merit in the social sciences and humanities, 1951–1971

Angela Tsay[a], Michèle Lamont[b,*], Andrew Abbott[a], Joshua Guetzkow[b]

[a]*University of Chicago, USA*
[b]*Department of Sociology, Princeton University, Wallace Building, Princeton, NJ 08540, USA*

Abstract

This paper investigates the questions of whether and how the evaluation of merit in academic disciplines changed between the early 1950s and the late 1960s. We analyze letters of recommendation written for prospective graduate students who applied to the Woodrow Wilson Fellowship Program during the periods 1951–1955 and 1967–1971, in the disciplines of economics, political science, philosophy, English and history. We find that in all disciplines, the relative use of intellectual and technical criteria increased during this time, the relative use of moral and social background criteria declined, while the use of personal criteria did not change. We find little evidence of disciplinary differences. In suggesting potential explanations for these findings, we focus on the impact of the dramatic growth and expanding diversity of academia during the postwar years.

1. Introduction

In this paper, we investigate the question, of whether and how the evaluation of merit in academic disciplines changed during the postwar era. Merit in academia is typically assumed to have become increasingly defined as academic ability during this period (Jencks and Riesman, 1968; Wilson, 1979). Except for undergraduate admissions (Synnott, 1979; Lemann, 1999), however, this assumption has not been empirically demonstrated. Moreover, while studies of disciplinary differences would lead one to expect the extent and timing of this change to vary across disciplines, this

* Corresponding author.
E-mail address: mlamont@princeton.edu (M. Lamont).

0304-422X/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0304-422X(03)00002-0

too has not been established. We explore and document changes in definitions of merit in academia by analyzing letters of recommendation written for prospective graduate students who applied to the Woodrow Wilson Fellowship Program at the beginning and end of the program's existence as a national fellowship competition: during the periods 1951–1955 and 1967–1971. We assess disciplinary differences by comparing letters written for applicants in the following disciplines: economics, English, history, philosophy, and political science. Our findings indicate that the relative use of intellectual and technical criteria increased substantially during this time, while the relative use of moral and social criteria declined. In addition, we find scant evidence of disciplinary differences.

Historical and social scientific accounts of changes in the US during the twentieth century indicate the increasing importance of criteria of merit based on academic ability rather than on social origins (see e.g. Jencks and Riesman, l968; Abbott, 1988; Bender, l997; Lemann, 1999). By the early 1950s, observers were already hailing the triumph of "universalistic" criteria in academia (Merton, 1949; Parsons, 1951).[1] Indeed, where else would one expect academic ability to be of utmost importance if not in academia?

But for one exception, social scientists have yet to study empirically how merit is defined in academia.[2] Lewis (1998) examined letters of recommendation written by faculty members for their colleagues and graduate students applying for academic jobs in the late 1960s and mid-1990s. Based on his findings, he challenged the notion that such universalistic criteria as intellect, competence and quality of work are the only considerations used by scholars to assess one another. He found that although academic ability and the quality of the scholar's work are most often mentioned in letters, other criteria, such as personality and character, are also discussed. However, he does not provide systematic and specific information on the relative use of academic criteria compared to extra-academic criteria. Moreover, although he analyzes differences between disciplines (considering chemistry, sociology and English), he does not study changes overtime within or between disciplines. We complement his work by addressing these issues and pursuing them further back in time by looking at the early 1950s. This is important, because accounts of the two decades following the Second World War mark it as the period when meritocracy—which places supreme importance on such universalistic qualities as intelligence and competence—became firmly entrenched in institutions of higher education as well as the national psyche (Lemann, 1999).[3] Thus, changes in definitions of merit are likely to have occurred since the 1950s.

[1] We recognize that even seemingly universalistic criteria, such as academic ability and test scores, may be biased in favor of individuals with privileged social backgrounds (Bourdieu, 1996; Lamont and Lareau, 1988; Jencks and Phillips, 1998; Fischer et al., 1996).

[2] The literature on the development of American academic disciplines in the post-World War II period tends to be descriptive and to consider disciplines largely independently of one another. For example, see *American Academic Culture in Transformation,* edited by Thomas Bender and Carl Schorske (1997).

[3] The term, "meritocracy", was only coined in 1958. See Michael Young's (1958) book, *The Rise of Meritocracy*.

We compare social science and the humanities disciplines to explore whether disciplinary differences in criteria of merit existed and whether universalistic criteria diffused at the same pace across these fields.[4] The postwar period saw the debate over the "two cultures," explicitly contrasting the scientific methods and culture of objectivity of the sciences with the normative and descriptive thinking of the older social sciences and humanities (Snow, 1959). Implied in this debate was the notion that universalism and meritocracy were not only more prevalent in scientifically-oriented disciplines, but would also penetrate faster in the sciences than in the humanities, as measurement of merit could be more objective and exact in the former than in the latter. In this context, the social sciences and the humanities could be seen as diverging, as the social sciences would increasingly borrow scientific norms—as well as a culture of universalism—from the sciences. Alternatively, the social sciences and the humanities could be seen as converging as a culture of universalism diffused throughout higher education, independently of the more normative vocation of humanistic fields. Factors promoting such a convergence include the increasing professionalization of academia, demographic pressures on institutions of higher education as a result of increasing size and diversity, and the expanded role played by government and private philanthropy in higher education during this time. Letters of recommendation written for Woodrow Wilson Fellows can be used to empirically assess these alternative hypotheses.

The paper tackles sequentially the twin processes of a general change in academia on the one hand and potential disciplinary differences on the other. We begin first by discussing the relationship between meritocracy and criteria of evaluation, suggesting why we expect letters of recommendation to become more universalistic through the postwar period, and then turn to the data and results bearing on this question. Only then do we discuss issues related to disciplinary divergence or convergence and, with new analytic techniques, assess the evidence for it.

2. Meritocracy and criteria of evaluation

Although any mention of merit today is virtually synonymous with academic ability, the criteria that constitute merit can change over time and differ culturally and institutionally (Brim et al., 1969; Friedland and Alford, 1991). As Karen (1985) notes, "abstracted from particular historical and sociological contexts, [merit] is a purely formal criterion. ... Thus, it is not merit as such, but the particular content of what is defined as meritorious and how it came to be defined that way that is at issue" (42–43).[5]

[4] The disciplines selected were established fields by the postwar period and cover the full range of disciplines eligible for the Wilson fellowships during the period 1951–1955, which excluded the natural sciences. Among the eligible disciplines, we selected fields that presented a broad range of variations across the social sciences and the humanities.

[5] As Sen (1999: 12–13) notes, such notions implicitly define merit as an innate quality of individuals, which is inherently worthy of reward, in contrast to a definition that Sen prefers, which is merit defined as bringing about some social good. In Sen's terms, our thesis is about changes in the set of intrinsic individual qualities that are valued.

Contemporary debates about "meritocracy" tend to focus on the legitimacy and consequences of using test scores as a gate-keeping mechanism for access to higher education (for example, see Arrow et al., 1999). While this work is certainly important, it leaves unexplored the conditions for the institutionalization of meritocracy: not just the widespread use of standardized tests, but also changes in the criteria deemed relevant, legitimate and/or legal for exercising gate-keeping discretion (though see Schudson, 1972; Synnott, 1979; Karen, 1985; Dobbin et al., 1993; Lemann, 1999). The creation of a meritocracy implies a fundamental cultural shift in the conceptions among institutional gatekeepers about what constitutes "merit"—from qualities of character (social and moral qualities) to academic ability, which comes to be seen as a more relevant and legitimate metric by which to measure people.

Whether such a shift occurred in the US has yet to be empirically demonstrated, but its timing would most likely have occurred in the first two decades after World War II. Historical studies indicate that between the first and second World Wars, college admissions became explicitly competitive as colleges began to place limits on enrollment (Wechsler, 1977). During these years, getting into a selective school had much more to do with having a good character (which was typically judged on the basis of one's social background) than it did with having top scores on entrance exams (Synnott, 1979). It was only after World War II that high school grades and S.A.T. scores came to occupy pride of place in determining undergraduate enrollment (Jencks and Riesman, 1968; Lemann, 1999).[6] We argue that the relative importance of academic excellence for entry to graduate school also increased during this time. Thus, we expect to find that, compared to letters of recommendation written in the early 1950s, letters written in the late 1960s will focus much more on intellectual and technical criteria as opposed to social, moral or personal criteria (see Appendix A for a description of these different criteria).

The prediction that universalistic criteria grew increasingly important for entry to graduate school during this time might seem so obvious as to need no empirical confirmation. However, there are good reasons to believe that no significant changes occurred over this period. To begin with, we might expect the criteria deemed relevant for graduate—as opposed to undergraduate—education to have developed and solidified around the turn of the century when graduate education became institutionalized in the US (Veysey, 1965). Since then, the primary purpose of graduate school—the education and training of experts, researchers and teachers—has changed very little (Berelson, 1960). Hence, it is quite conceivable that intellectual or "universal" criteria trumped other criteria (such as social or moral ones) as a component of the institutional logic of academia from the beginning of the century (Friedland and Alford, 1991), and that these criteria simply spread to undergraduate education in the postwar period. Alternatively, if any substantial change did occur, it might have taken place during the interwar years as academia became more professionalized (Berelson, 1960). In either case, we would not expect to find evidence of any changes in the relative importance of different criteria of evaluation in the

[6] In the section below on disciplinary convergence, we suggest causal factors that are also relevant for understanding why this change occurred when it did.

postwar period. While our data do not permit us to assess the level of change across periods, they can establish the occurrence of change in the postwar era, which is in itself an important contribution given the paucity of our knowledge concerning academic criteria of evaluation.

3. Data and measures

In the course of its operation as a national fellowship competition, the Woodrow Wilson Fellowship program designated over 15,000 college seniors as Fellows. By providing funding for the first year of graduate schooling, the program sought to encourage these students to pursue careers in academia. The enterprise was phenomenally successful—Woodrow Wilson Fellows entered the faculties of virtually every US and Canadian institution of higher learning (Rosenhaupt, 1969). Many have been influential both in their fields and in society more generally: for example, Gary Becker in economics, James Q. Wilson in political science, Robert Pippin in philosophy, A. Bartlett Giamatti in English, and William H. Sewell, Jr. in History.

These Fellows represent an elite cadre of scholars who went through several levels of selection before becoming fellowship recipients: they were nominated by professors from their undergraduate institutions, evaluated by regional and national selection committees on the basis of their academic records, recommendations and personal essays, interviewed by committees of college professors from their regions, and designated as Fellows by national selection committees. We regard the letters written on behalf of these candidates as an exceptionally valuable source of information about the shifting valuation of merit in academic disciplines, both because of the prestige attached to the fellowships, and because of the continuity in the management of the program over time.[7]

We coded recommendation letters contained in a sample of Fellows' files for the periods 1951–1955 and 1967–1971. Four-year time spans were selected in order to minimize minor period effects and to ensure adequate sample sizes. To ensure the comparability of the two sets of data, we sampled only files of male, non-minority applicants from US undergraduate institutions.[8] We coded only those

[7] Note that we do not consider letters written for unsuccessful applicants since these were not preserved by the Woodrow Wilson Foundation. We have no reason to believe that the criteria of evaluation used in the letters written for successful and unsuccessful candidates varied. The absence of consideration of unsuccessful applicants poses a problem for our analyses only if success was correlated with the evaluation repertoires of letters in either time period. The standardized and more or less stable wording of the recommendation letter prompts (see below) argues against this possibility. Moreover, both letter-writers and Fellowship juries came from the same disciplinary populations, even if the latter were somewhat more elite. If juries differentially rejected candidates with more "personal" recommendations, it would imply that juries had different expectations from at least one subgroup (the more "personal" letter-writers), an implication that seems unlikely given the common disciplinary origin.

[8] While it would be interesting to investigate the ways in which recommendation letters written for minorities and women differ from those written for white men, insufficient data preclude such an analysis. Although there were 200 fellowships awarded annually during the period 1951–1955, only a tiny fraction were held by minorities or women (there was an official quota limiting the number of women recipients until 1958).

recommendations for which the discipline of the recommender was the same as the destination discipline of the applicant, since this enables us to see disciplinary differences in the purest form possible, considering that the review committees were multi-disciplinary.[9]

Stratified random sampling was employed to increase the representativeness of the sample while decreasing sampling error (Babbie, 1989). This initial procedure was then supplemented by disproportionate sampling within discipline and time strata to gain sufficient numbers of cases from each sub-population of interest (Kish, 1965). For the period 1951–1955, a directory of Fellows published by the Woodrow Wilson National Fellowship Foundation was used to establish the sampling frame from which 20 applicant files from each of the five disciplines were randomly selected. For the period 1967–1971, the sampling frame was established by constructing a partial catalog of the relevant holdings of the Foundation. Again, 20 applicant files from each of the five disciplines were randomly selected. Applicant files which did not fall into the delimited definition of the population were discarded and recommendation letters from the remaining files fulfilling the criteria outlined above were photocopied for coding. We coded a total of 208 letters of recommendation. For economics, political science, philosophy, English, and history, there were 16, 13, 22, 22, and 28 letters respectively for the period 1951–1955, and 22, 18, 20, 27, and 20 letters respectively for the period 1967–1971.

The instructions given to letter-writers differed only slightly in the two time periods. For the period 1951–1955, recommenders were asked to evaluate whether

> "on the basis of your experience and measured on the triple criteria of intellect, character and personality, the individual shows high promise for the career of teaching and scholarship ...[and to assess the candidate's] ability to carry on advanced studies in his field and achieve a successful career in it."

During the period 1967–1971, recommenders were asked for their

> "opinion of the candidate's intellect and character, particularly those qualities which bear on his or her promise as a college teacher and scholar...[and to] describe the student's qualities of mind and personality, particularly the state of his preparation for graduate study and his ability in the writing of reports on independent work."[10]

[9] One might think that the peculiar nature of the Wilson competition would militate against this focus on intra-disciplinary intellectual matters. Since the Fellowship juries were multi-disciplinary, one might expect a disciplinary-neutral vocabulary. Indeed, a vocabulary of personal and character qualities would have fit such a model. But it is striking that the word "interdisciplinary" did not occur even once in any of the letters read. So far as we could tell, disciplinary referees characterized their candidates in disciplinary terms, even for a multi-disciplinary panel.

[10] Although the wording of the prompts is not identical for the two periods, both ask specifically for information on intellect, character, and personality in reference to scholarship and teaching. Thus, there is no differential emphasis placed on these types of criteria across the two periods. Moreover, we discussed the issue with Judith Pinch, vice president of the Woodrow Wilson National Fellowship Foundation, who

The letters of recommendation were coded in their entirety. A characteristic letter can be found in Appendix B. Each descriptive word used by a letter writer received a separate code. In most cases, assignment of codes was unproblematic, being essentially a matter of listing exact words. For example, the adjective "logical" received the code "systematic/logical/analytic".[11]

The entire set of letters was coded twice into micro code categories, which were then grouped into 20 meso categories:

(1) general academic ability
(2) critical/analytic ability
(3) interpretive ability
(4) other academic ability
(5) academic communication
(6) intellectual desire
(7) universal intellectual criteria
(8) culture and cultivation
(9) class markers
(10) social graces
(11) physical description
(12) extra-curriculars
(13) moral character
(14) work ethic
(15) social consciousness
(16) psychological/emotional health
(17) personality
(18) interpersonal relations
(19) organizational skills
(20) technical skills

These 20 clusters of descriptors were then further organized into five macro categories, which we have denoted: (1) intellectual (intellectual capacity, meso categories 1–7), (2) technical (skills, meso categories 19 and 20), (3) social (social attractiveness, meso categories 8–12), (4) moral (moral character, meso categories 13–15), and (5) personal (psyche and personality, meso categories 16–18). Criteria of intellect have to do with intellectual powers or drive and are exemplified by descriptors such as "insightful," "promising," and "good academic record". Technical criteria have to do with technical abilities or skills, such as being "efficient", or having "good foreign language aptitude" or "knowledge of statistics". Social criteria refer to social attractiveness; examples are "cultured," "poised," or "attractive". Moral criteria have to do with moral character and commitment and are epitomized by qualities such as being "loyal" or "diligent." Finally, personal criteria have to do with

was involved with the program between 1959 and the time it ended in the early seventies. She confirmed our belief that there was no change within the foundation in the emphasis placed on teaching or research during the course of the fellowship program, though a foundation history of the program (George, 1967) indicates that having a strong motivation towards remaining in academia was perhaps given more importance in later years. Hence we have no reason to believe that changes in the priority of the Woodrow Wilson Foundation fellowship program can account for our findings.

[11] The coding scheme and code categories were developed over several stages of pre-testing. First, in an effort to derive codes semi-inductively, a sub-sample of the letters was coded. Several broad groups of categories emerged from this trial coding. With these broader categories in mind, additional codes were added to the initial scheme. Code categories were finalized after repeated waves of testing and modification. The final coding was performed by only one of the authors. However, the entire set of letters was coded twice, and discrepancies settled by returning to the text of the letters. Given the transparency of the original coding decisions, a second full coding was judged unnecessary.

personality and psyche; examples are "enthusiastic", "shy," and "witty". (See Appendix A for a more detailed break-down of the macro categories.) Of these criteria, we call intellectual and technical criteria "universalistic" and social, moral and personal criteria "particularistic."

4. Results

Fig. 1a–e shows the per-letter mean number of uses of intellectual, technical social, moral, and personal evaluative criteria for the disciplines economics, political science, philosophy, English, and history in the two time periods. Thus, intellectual criteria were mentioned a little under six times per letter in English in 1950 and slightly under 10 times per letter in the same discipline in 1970. These means are very

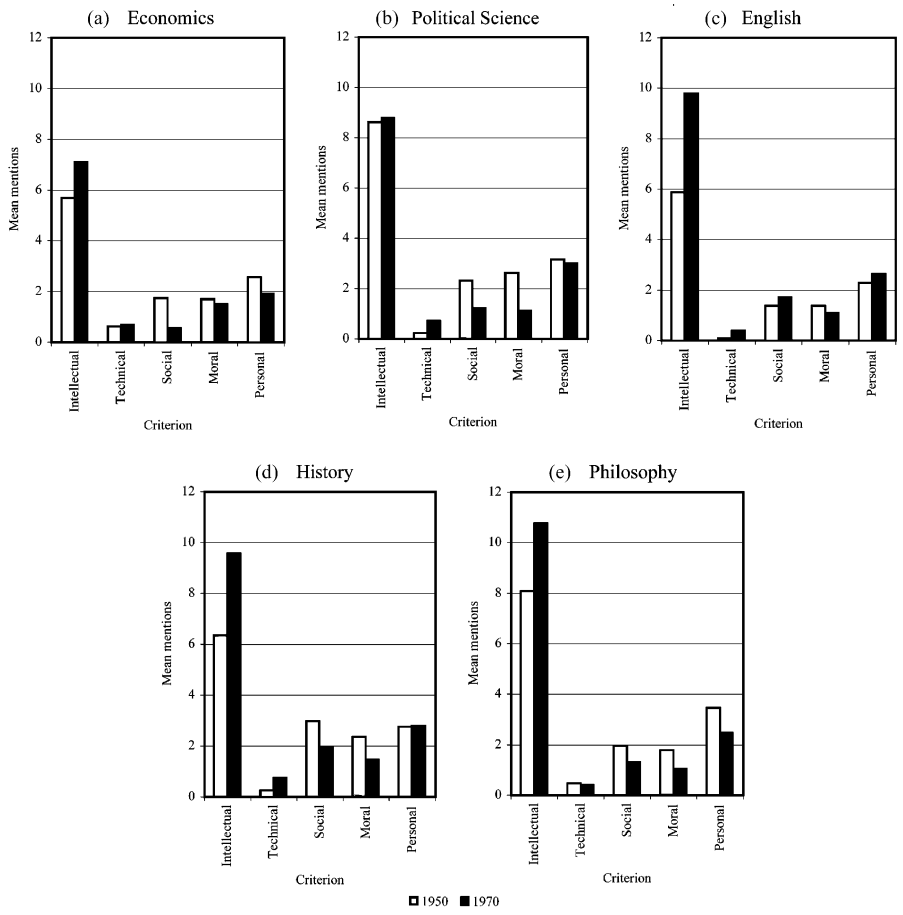

Fig. 1. (a–e) Mean usage of criteria of evaluation, by discipline.

unstable. For all but the intellectual criteria, standard deviations are of the same size, roughly, as the means.

The figure shows an obvious increase between 1951 and 1971 of the mean usage of intellectual and technical criteria for all disciplines (except for technical criteria in the case of philosophy). The opposite trend obtains for social and moral criteria in all disciplines (except for social criteria in the case of English). Personal criteria change inconsistently; per letter use of personal criteria decreases for economics, political science, and philosophy, but increases for English and history.

Tests for the statistical significance of change over time require that the data be approximately normally distributed. As the large standard deviations imply, this assumption is met only by the distributions of intellectual criteria, so we limit *t*-tests of change to the use of intellectual criteria.[12]

These tests indicate that the use of intellectual criteria as measured by average mentions per letter is significantly greater in the later period than in the earlier, but only for the disciplines of English and history. The change in philosophy is borderline. All five disciplines do however change in the intellectual direction. There is thus some evidence that recommenders in the period 1967–1971 made more reference to intellectual attributes than did their counterparts in the period 1951–1955.

This apparent move toward intellectual criteria could simply reflect longer letters overall. Since the other criteria are relatively scarce, they could have expanded proportionately without achieving statistical significance. Recasting the data as proportions controls for this possibility. In this view, we consider, for each letter, the *proportion* of that letter's total descriptors that are intellectual, technical, and so on. Descriptively, over the period 1951–1971 *relative* (i.e. proportional) reliance on intellectual and technical criteria increased, while relative reliance on social and moral criteria decreased, and relative reliance on personal criteria remained basically stable.[13] These descriptive results regarding proportional usage (or, relative salience) are consistent with those found above when looking at the raw data.

As before, statistical tests for changes over time are invalid except for the category of intellectual criteria because they are the only criterion type to be approximately normally distributed. As shown in the right-hand panel of Table 1, the proportional use of intellectual criteria shows a statistically significant increase between 1951 and 1971 across all disciplines except political science. Note, too, the dramatic increase in the relative importance of intellectual criteria during this period. Intellectual criteria went from being half to two-thirds of all criteria mentioned in philosophy. In economics and history, whereas intellectual considerations did not constitute a majority of the criteria used in the letters of recommendation in the first period, these criteria represented a large majority in the second. These results provide reinforcing evidence of significant change over time in the evaluation of merit: an increase in the relative importance of intellectual criteria compared to all other criteria.

[12] We test for differences in the other criteria using Poisson models in the next section of the paper.

[13] Because of space limitations, the equivalents to Fig. 1a–e are not shown but are available upon request.

Table 1
Means, proportions, and statistical tests of change of intellectual criteria, 1951–1971

	Means			Proportions		
Discipline	μ 1950	μ 1970	Pr > | t |	π 1950	π 1970	Pr > | t |
Economics	5.69	7.09	0.23	0.46	0.6	0.05
Political Science	8.62	8.78	0.92	0.51	0.59	0.19
English	5.86	9.78	0.01	0.54	0.63	0.04
History	6.36	9.55	0.04	0.43	0.58	0.02
Philosophy	8.09	10.75	0.09	0.51	0.67	0.01

5. Disciplinary differentiation

The relative stability of the results across disciplines should not, however, blind us to the possibility of disciplinary differentiation. As suggested in the introduction, there are good reasons to expect divergence and convergence between disciplines over this time.

5.1. *Difference and divergence*

Both through random drift and through the deliberate construction of boundaries (Gieryn, 1999), we would expect the differentiation of academia into relatively autonomous departments to result in differences in the types and qualities of work performed in these disciplines. As a result, distinct reputational communities should have formed collectivities with their own communication systems and evaluation criteria (Whitley, 1984). Competition between disciplines over material resources and cultural jurisdiction would then promote the elaboration of further disciplinary differences as academic professions sought to clearly distinguish themselves from one another (Ross, 1990; Veysey, 1965; Larson, 1977; Abbott, 1988; Gieryn, 1999).

It would seem reasonable, then, to expect that in writing recommendation letters for these Fellows, recommenders drew upon the distinct cultural values and evaluative criteria of their disciplines (see for example Bourdieu, 1988; Becher, 1989). Differences between disciplines in the relative salience of various evaluative criteria would then be evidence for broader differences among disciplinary cultures of the evaluation of merit between the first and second periods. Snow's (1959) essay on the fault lines dividing scientists and humanists suggests that differences and divergence are likely to be found between the humanistic and social scientific disciplines we study. Whitley's (1984, esp. chapters 3 and 4) framework suggests a specific hypothesis: Compared to humanistic disciplines, scientific ones tend to have higher levels of what he calls strategic and functional dependence and lower levels of task uncertainty than humanistic ones. Consequently, standards of quality assessment tend to be more agreed upon, and standards of what constitutes intellectual and technical ability are more widely shared in scientific disciplines as compared to humanistic ones. Building on Whitley, we would expect universalistic criteria to

diffuse more slowly among humanistic disciplines. Of the disciplines included in this study, we expect intellectual and technical criteria to be of greater importance in the social science disciplines (in this case, economics and political science) than in English, history and philosophy.[14]

5.2. Similarity and convergence

We also have reason to expect little or no differences between disciplines in the relative importance of criteria, because a similar set of forces was operating on higher education as a whole that promoted the increasing importance of universalistic criteria across all disciplines. At the most general level, merit defined as academic ability and certified by educational credentials can be seen as part of the larger struggle between upper-class elites and the rising middle class trying to legitimize academic ability and educational credentials as bases for access to resources. In this context, universal criteria, and intellectual merit in particular, became a primary component of the particular means by which power and privilege came to be legitimized (Veysey, 1965; Oleson and Voss, 1979; Collins, 1979; Abbott, 1988; Bourdieu, 1988, 1996). These changes went hand in hand with a decline of emphasis on character.[15]

Important transformations accentuated these shifts. The sudden and immense population growth and expansion of diversity in academia during the postwar era probably had much to do with changes in criteria of evaluation.[16] Paralleling these changes in the student population—and even more important—were the expansion, diversification, and cohort replacement of the professoriate. The passage of time and a tremendous increase in the size of the university system meant that only a small proportion of the professors writing letters during the 1967–1971 period were left over from the early 1950s. By the late 1960s academia was dominated by an entirely new generation of professors, who were educated in the postwar era (American Council on Education, 1970). These changes could have affected the definitions of merit in a few ways.

[14] Lewis (1998) found only limited evidence for differences between disciplines in the weight given to certain criteria. The only notable exception is the relative unimportance of interpersonal skills (a component of "personal" criteria in our classification) in English, compared with sociology and chemistry (pp. 63–76).

[15] For instance, expertism was adopted as a legitimating ideology by British apothecaries because character was already claimed as a legitimating ideology by their opponents, the physicians, and not because there was any serious evidence that apothecaries were in fact more expert. Later in the 19th century, the replacement of character by expertise seems to have been driven more by the breakup of traditional elite networks, which meant that character credentials were harder to evaluate. Personal character knowledge was communicable over much shorter social distances than were expert credentials. That was indeed the whole point of credentials as opposed to personal testimony (Abbott, 1988).

[16] Funded by the G.I. Bill, over three million students attended some form of post-secondary education in the immediate postwar period (DeVane, 1965). The G.I. Bill had strong leveling effects; families across the socioeconomic spectrum began to send their children to college (Jencks and Riesman, 1968). Jews and Catholics, and later minorities and women, took their place in the ranks of academia as never before (Berelson, 1960; Synnott, 1979).

First, paralleling the turn of the century conflict over legitimacy between old and new elites, the more explicit use of social background as a gatekeeping criterion ceased to be legitimate, especially as access to higher education became more open during this time. Cohort replacement of faculty is especially important in this respect, since the middle class was better represented among the new generation of professors. For example, in 1955, 23% of professors in the social sciences had fathers with professional occupations other than teaching and 15% had fathers who were manual laborers (Lazarsfeld and Theilens, 1958: 7); by 1968, this had reversed to 14% and 24% respectively (Ladd and Lipset, 1975: 172–3; also see Berelson (1960) and Harris (1972) on the changing class composition of the professorate).

At the same time, in the suddenly larger and more diverse higher education community, personal qualities like integrity and wit were not as mutually recognizable as they would have been in academic disciplines that previously numbered perhaps a few thousand nationwide.[17] Diversity meant mutual incomprehensibility—both between faculty and students and between faculty and faculty. On this account, universalistic criteria helped to create a commonly accepted language in a newly cacophonous world.

Complementing these factors, the rapidly increasing size of the university system pushed academia toward hierarchy, promoting objective criteria and expertise as ways of coping with coordination problems (Collins, 1979; Whitley, 1984). Increased size also made academia a much broader market, in which objective criteria may have been valued because they made decision-making more efficient and therefore less contestable (Oleson and Voss, 1979). Finally, the increasing dependence of academia on outside funding (both governmental and philanthropic) promoted a move towards standardization of evaluative criteria (Hofstadter, 1961; DeVane, 1965; Geiger, 1993). This trend had strong affinities with a rhetoric of efficiency and standardization that fit better with universalistic criteria of evaluation than with social or moral ones. With all these forces operating on higher education more or less as a whole, we would expect similarity between disciplines in the relative importance of intellectual and technical criteria and a convergence between disciplines in the importance of these criteria from the earlier period to the later one.

6. Results

Descriptive analysis of data using multidimensional scaling of the mean usage of the five types of criteria by discipline in each time period did not reveal disciplinary differences in the criteria used, although it did suggest a uniform change across dis-

[17] Between 1940 and 1970, the number of faculty employed in colleges and universities in the US increased from 150,000 to 600,000—with a third of this growth coming between 1965 and 1970 (Ladd and Lipset, 1975: 169). During this period, the religious background of newly minted professors was changing as well. In 1955, two thirds of professors in the social sciences identified themselves as protestant, with only 50 doing so in 1968. Catholics and Jews also made substantial inroads into academic teaching during this period (Ladd and Lipset, 1975: 170).

ciplines between the first and second period.[18] In order to test this more rigorously, we model raw counts of the usage of each criterion per discipline using Poisson regression.[19] These models predict the number of mentions of a particular evaluation criterion in the *i*th letter as a function of a base rate, a time variable (first wave, second wave), and a discipline variable. The estimates are produced by maximum likelihood methods by STATA (release 5).[20] In order to ensure stability of the calculations, we chose reference categories for the model so that each type of criterion would be stably and non-degenerately represented in the reference cells. Of several possible choices, we used political science and the second time period. Where the Poisson models do not fit, we undertake similar negative binomial regressions, fitting the letter-writer heterogeneity directly.[21] Results of these models are shown in Table 2.

The first column of results is from a Poisson regression of counts of technical criteria on discipline and time. The remaining columns present data from negative binomial regressions. Unlike our previous analysis, here we are able to assess changes independently for each type of criteria. As shown in Table 2, there are no disciplinary differences over time in the use of any of the five types of criteria, with the lone exception of economics in the case of intellectual criteria. There are, however, strong, statistically significant time effects, all in directions consistent with results presented earlier. Being located in time period 1 is negatively correlated with counts of technical and intellectual criteria—which means that the use of technical and

[18] The only disciplinary difference it did suggest was in the length of letters, but in a separate OLS regression predicting letter length, no disciplinary difference were found, largely as a result of large individual variation in letter length. These results, as well as those of the multidimensional scaling, are available upon request.

[19] The intense variability in criteria used implied by small means and large variances of the evaluation criteria distributions summarized in Fig. 1 suggested that the number of letters with N mentions of a particular criterion will obey not the normal distribution, but the Poisson. This was easily tested. There are five disciplines, two periods, and four rare criterion types—technical, social, personal, moral (recall that intellectual criterion scores are normally distributed. Of the 40 such combinations, six have so little data that distribution fit cannot be tested. But of the 34 remaining, only five rejected the Poisson distribution. (These five were not concentrated in any one discipline, time period, or criterion.) This implies that letter-writers use the various criteria randomly, at rates that may vary by discipline and time. Only in the five rejections did we see clear evidence of underlying individual heterogeneity in letter-writers' propensity to use certain criteria. If this was not the case, then the combined distribution across the whole discipline would not be Poisson, but negative binomial. The negative binomial distribution would also arise if there were contagion effects between letter-writers. However, the study design makes these nearly impossible.

[20] Note that doing a separate model for each criterion assumes that the number of mentions of each criterion type is independent of the others. Against this assumption, one might argue that writers whose letters have become long might reduce the personal comments, which normally come at the end, or writers may add personal comments only if they feel a need to fill out a letter light in intellectual meat. But this seems a minor concern.

[21] Note that under the regression format we run one model for each criterion and thus all the individual heterogeneity from ten separate discipline/period combinations is combined in them. Hence, it should not be surprising that for all but the technical criteria, the Poisson is rejected (as opposed to its acceptance in the earlier separate fits) and we are driven to negative binomial regression. We fit the intellectual criterion as well. Although it was not necessary to use a Poisson approach in the case of intellectual criterion, the Poisson can be used to approximate binomial processes in parameter regions where we would usually use the normal. See Feller (1957).

Table 2
Results from Poisson and negative binomial regressions, by type of criterion[a]

Independent variable	Poisson regression	Negative binomial regression			
	Technical	Intellectual	Social	Moral	Personal
Time	−0.59*	−0.31*	0.40*	0.43*	0.09
	(0.22)	(0.08)	(0.15)	(0.13)	(0.11)
Economics	0.24	−0.30*	−0.48	−0.08	−0.34
	(0.32)	(0.13)	(0.28)	(0.22)	(0.19)
English	−0.73	−0.09	−0.05	−0.37	−0.22
	(0.38)	(0.12)	(0.25)	(0.22)	(0.18)
History	−0.03	−0.09	0.36	0.07	−0.12
	(0.33)	(0.12)	(0.24)	(0.20)	(0.18)
Philosophy	−0.13	0.09	−0.05	−0.23	−0.04
	(0.34)	(0.12)	(0.26)	(0.22)	(0.18)

[a] Standard errors are in parentheses.
* $P<0.01$.

intellectual criteria increases between 1951 and 1971. Also consistent with our prior findings is that the use of social and moral criteria decreases between the two time periods, and no change is found for personal criteria.

Since these results for raw counts could be confounded with length effects, we recast the data in terms of proportions.[22] Transforming the raw count data into proportional counts allows us not only to assess relative salience (and to divide out the effect of length of letters), but it also makes it feasible to run one comprehensive negative binomial regression model with discipline, time, and type of criterion as predictors.[23]

Our final analyses are negative binomial regressions of proportional counts on discipline and criterion type for each of the time periods separately. We then include data for both time periods simultaneously and model count as a function of discipline, criterion type, and time. The reference categories for discipline, criterion type, and time are political science, moral, and the 1967–1971 period, respectively.[24]

The first two columns of Table 3 present the negative binomial regressions of proportional counts on discipline and type of criterion. In both the model for 1951–

[22] To do this, we divided each count by its row total (the total number of mentions in that letter), multiplied by 100, and rounded to the nearest integer. Thus, the proportions of intellectual, technical, social, moral, and personal criteria sum to 100 for each letter. To distinguish these converted count data from raw counts, we refer to them as proportional counts.

[23] With the data as raw counts, models had to be run separately for each type of criterion since we would not have been able to take into account the intra-letter correlation across the five types of criteria that was induced by length differences if we had run a comprehensive model with criterion type included as a predictor.

[24] Note that all three models exclude data on personal criteria. When the raw count data are transformed into proportional counts, they sum necessarily to 100, and knowing four of five criterion proportions for a given discipline at a given time is equivalent to knowing the proportions for all five criteria. We chose to treat personal criteria as the "missing" value because our previous results showed the personal criterion to undergo the most contradictory changes over time.

Table 3
Results from negative binomial regressions using proportional counts[a]

Independent variable	Period-specific models		Pooled model
	1951–1955	1967–1971	
Time			0.06
			(0.09)
Economics	0.30	0.18	0.19
	(0.21)	(0.22)	(0.16)
English	−0.11	−0.17	−0.17
	(0.20)	(0.21)	(0.15)
History	0.15	0.13	0.12
	(0.19)	(0.22)	(0.15)
Philosophy	0.01	−0.12	−0.10
	(0.20)	(0.22)	(0.15)
Intellectual	1.24*	2.13*	10.66*
	(0.15)	(0.19)	(0.13)
Technical	−2.02*	−0.63*	−10.28*
	(0.17)	(0.19)	(0.13)
Social	−.06	0.22	0.05
	(0.16)	(0.19)	(0.13)

[a] Standard errors are in parentheses.
* $P<0.01$.

1955 and the model for 1967–1971, the only statistically significant results are for the coefficients on intellectual and technical criteria. Relative to moral criteria and controlling for time and discipline, intellectual criteria are employed in greater proportions, whereas technical criteria are employed in smaller proportions.[25]

Our main predictor of interest in this section is not, however, criterion type, but discipline. None of the coefficients on the dummy variables indicating discipline is significant in either of the two models. This tells us, again, that there are no significant disciplinary differences in the *proportional* use of criteria of evaluation during either the period 1951–1955 or 1967–1971.

Results from our final model are presented in the final column of Table 3. This model pools data from the two time periods and performs a negative binomial regression of proportional count on discipline, type of criterion, and time. Again, only the coefficients on intellectual and technical are significant, and there are no disciplinary differences whatsoever.[26]

[25] The contradiction of the technical finding with the mean figures in Fig. 1 should not be surprising. Technical comments were very rare and hence results are likely to be unstable. It is probable that technical criteria was not a useful category, and that technical criteria should have been lumped in with intellectual.

[26] The coefficient on time is also not significant, but this is expected since the test for the significance of time is in effect a test of whether there was change over time in the "missing" (as opposed to reference) category, in this case personal criterion. As we showed earlier, in contrast to change in the relative salience of intellectual, technical, social, and moral criteria, there is no statistically significant change over time in the relative salience of personal criteria (Table 3). The reader should note that the criteria effects are relative to the omitted category (moral).

In summary, our results do not support the existence of disciplinary differences in the evaluation of merit. We found disciplinary differences neither in the mean use of criteria nor in the relative salience of different types of criteria. There were no disciplinary differences within and between time periods. These results also provide further evidence that the relative use of intellectual criteria increased between the first and second time periods.

7. Conclusion

All of our analyses indicate that the evaluation of merit changed substantially during the postwar era. Between the early 1950s and late 1960s, the relative use of intellectual criteria clearly rose and technical criteria very likely rose, while use of moral and social criteria declined. The use of personal criteria did not appear to change.[27] These findings hold up from the most descriptive to the most technical analyses. We found no differences between disciplines during and between either period.

We argue that these changing criteria represent shifts occurring more broadly in the social sciences and humanities. There are, however, two alternative explanations that would suggest a coincidental relationship between the changes we document and the growth of meritocracy:

(1) The change could simply be a product of the increasing competitiveness of graduate school overall and the Wilson Fellowships in particular. This increased competitiveness would lead letter writers to inflate their recommendations by saying more about the candidate and going into greater detail, which would translate into higher raw counts for period 2, as opposed to period 1. But, this explanation is not plausible, because we showed an increase in the *relative* use of universalistic criteria compared to particularistic ones. Moreover, even if the changes were simply a matter of "recommendation inflation", this would not explain why the letter writers inflated only intellectual criteria, unless it was precisely these criteria that had become more important by the late 1960s.

(2) Changes could result from the ephemeral political climate of the early 1950s, as opposed to more profound changes. This explanation would posit that assessments of morality and character were more salient during those years as a result of the long shadow that McCarthyism cast over academia. The late 1960s, however, were also politically sensitive, and there is just as good reason to suppose that faculty concern with subversive students during this time would lead them to place greater importance on moral and social criteria.

The findings are consistent with historical studies that identify the postwar period as the time when academic achievement and ability replaced social background as the primary criteria for determining undergraduate enrollment (Jencks and Riesman,

[27] Most of the extra-intellectual considerations that Lewis (1998) found in abundance in letters fall under this criterion.

1968; Wechsler, 1977; Synnott, 1979; Lemann, 1999). The notion that criteria other than academic excellence were at one time at least as important for graduate study may come as a surprise to many. We do not want, however, to suggest that intellectual criteria were ever unimportant; simply that relative to other concerns, the importance or relevance of intellectual criteria grew to become the dominant factor over the course of this study (as indicated by the increasing proportion of the letters devoted to discussing intellectual qualities).

The near uniformity of the changes we found across disciplines is consistent with our discussion of forces acting upon higher education that would lead to more reliance on intellectual criteria in all disciplines. Thus, conflict between elites, changes in the size and demographic diversity of the university system and an increasing dependence of academia on outside funding all seem to have potentially led to these changes in criteria of evaluation.

An important limitation to our study is that we have no data to indicate whether the importance that selection committee members placed on different types of criteria also shifted in the same way as the recommenders. Our argument about the institutionalization of new definitions of merit implies that this ought to have happened. We believe it reasonable to assume that, since writers and readers were taken more or less from the same population, changes in the criteria that recommenders emphasized mirrored changes in that criteria that were considered more important or relevant in their discipline and the world of higher education more generally.[28]

We also do not know the extent to which the review process was 'universalistic' in other terms: for example, whether candidates from more elite, religious or private universities were favored; or whether letters from more prestigious scholars were given more weight.[29] With academia expanding during this time period, committee members would have been less likely to know or be familiar with the recommenders; as a result, they may have relied more on institutional and personal prestige than on personal familiarity to judge the 'weight' of a given letter. We also cannot say anything conclusive about changes in racial and gender bias, although we believe that our findings about the decreasing relevance of what we define as social criteria would also extend to the race and sex of the applicant. On the other hand, due to the growing role of affirmative action during the late 1960s, it is possible that recommenders would have highlighted the identities of minority and women candidates, with good intentions. Although we cannot address such issues with the data that we have, these questions complicate the picture of the uniform changes we paint here and suggest lines for further study.

Future research might also investigate into the impact of the demographic and institutional factors that we outline[30] and delve into disciplinary differences at a finer

[28] We also assume more generally that emphasis on certain types of criteria, as measured by their proportional representation in letters, correlates with their relative weight in the decision-making process. This is not necessarily the case. If every candidate has a letter filled mostly with glowing things written about their academic ability (or alternatively, their character), then it may be the smaller emphasis placed on other details about them that "put them over the top" and hence, paradoxically, end up being more important.

[29] We thank an anonymous reviewer for pointing this out.

[30] See Tittle and Rotolo (2000) for an exemplary study along these lines.

level of detail, because differences between disciplines might very well operate at a level "below" the categories we identify. For example, a student in philosophy might be more likely to be hailed as brilliant because of her ability for meticulous analysis, whereas an English student might be considered brilliant because of his ability to convincingly integrate ideas with stylistic elegance. At our level of analysis, both students are judged based on intellectual criteria. An observational study of how intra-disciplinary and interdisciplinary peer review panels draw the boundary between worthy and less worthy proposals could also provide a more fine-grained understanding of definitions of excellence as they are articulated today (Lamont, 2000). Finally, interviews with fellowship applicants on their presentation of self and their understanding of the types of excellence valued by various fellowship programs could provide an important complement to how academic excellence is constructed over time and across settings (Lamont et al., 2000).

Acknowledgements

We thank Judith Pinch and the Woodrow Wilson Foundation for providing access to the data as well as an anonymous reviewer. A previous version of this paper was presented at the annual meeting of the American Sociological Association, Toronto, Canada (August 1997).

Appendix A. Descriptions and groupings of code categories

I. INTELLECTUAL: having to do with capabilities of the intellect and intellectual drive

Composed of 7 meso categories: (1) general academic ability; (2) critical/analytic ability; (3) interpretive ability; (4) other academic ability; (5) academic communication; (6) intellectual desire; and (7) universal criteria of intellect.

A. General academic ability

i. prepared
ii. promising
iii. genius
iv. gifted
v. best
vi. excellent

B. Critical/analytic ability

i. subtle/nuanced mind
ii. integrates/assimilates/synthesizes
iii. gets to the heart of the matter

iv. links concepts and data
v. good at abstraction/theory
vi. keenness of mind/critical
vii. insight/perception
viii. agile/flexible mind/thought
ix. systematic/logical/analytic
x. precise/meticulous
xi. objective
xii. good concentration

C. Interpretive ability

i. good at explication of texts
ii. deep/sophisticated understanding
iii. imaginative/original (work)
iv. not (very)imaginative/creative
v. comprehension with thoughtfulness
vi. not very sensitive reader

D. Other academic ability

i. not just cleverness
ii. careful/deliberate study
iii. command/mastery of the material
iv. mature work/understanding
v. thorough/close study/work
vi. thinks clearly
vii. independent work or research
viii. independent thinking
ix. good scholarship/scholarly
x. good memory
xi. good coursework or work (general)

E. Academic communication

i. expresses himself forcefully
ii. good expression of thoughts
iii. does not express himself well
iv. concise expression
v. well-written work/writes well
vi. articulate/well-spoken
vii. writes with style
viii. good class discussion
ix. good class presentations
x. excellent/good written work

F. Intellectual desire

i. calling
ii. sure about going to grad school
iii. not sure
iv. intellectual curiosity
v. commitment to academia/desire to learn
vi. enjoys his work
vii. pursues ideas tenaciously
viii. goes beyond requirements
ix. challenges himself
x. not materialistically driven
xi. alert/exhibits interest

G. Universal criteria

i. good academic record
ii. mention of academic record
iii. grades or performance (w/r/t undergrads)
iv. performance (w/r/t graduate students)
v. grade point average/class rank
vi. Phi Beta Kappa or Rhodes

II. TECHNICAL: having to do with technical abilities or skills that are not necessarily tied to intellect.

Composed of two meso categories: (1) organizational skills; and (2) technical skills.

A. Organizational skills

i. problems meeting deadlines
ii. lack of discipline
iii. good class attendance
iv. extremely well-organized mind
v. papers not well-organized
vi. well-supported papers and thought
vii. well-organized papers/work
viii. well-organized oral presentation
ix. needs greater discipline in writing
x. meets deadlines/punctual
xi. well-organized
xii. efficient
xiii. good work habits

B. Technical skills

i. foreign language
ii. no linguistic aptitude
iii. statistics/math/computers
iv. technical competence
v. skill

III. SOCIAL: having to do with social attractiveness

Composed of five meso categories: (1) culture and intellectual cultivation; (2) class markers; (3) social graces; (4) physical description; and (5) extra-curricular activities.

A. Culture and intellectual cultivation

i. wide/broad interests/eclectic
ii. erudition/breadth of knowledge
iii. liberal education
iv. narrow interests/not widely-read
v. specialized
vi. intellectual
vii. well-read/reads widely
viii. comments narrow in scope
ix. depth rather than breadth
x. cultured
xi. poise
xii. charming
xiii. literary sophistication
xiv. discriminating

B. Class markers

i. wide (life) experiences
ii. gentleman
iii. mention of family background

C. Social graces

i. tactful
ii. tactless
iii. at-ease socially
iv. good conversationalist

v. has presence/aura/popular
vi. first-rate fellow
vii. well-liked/respected by peers
viii. well-liked/respected by faculty
ix. extroverted/outgoing
x. not at-ease socially
xi. manners

D. Physical description

i. attractive/pleasing appearance
ii. clean-cut/well-groomed
iii. virile
iv. rugged (physical appearance)
v. strong/outstanding physique
vi. healthy
vii. physical stamina
viii. physical deformity
ix. dresses well

E. Extra-curricular activities

well-rounded
leadership qualities/student leader
not a bookworm or grind
square
women
extra-curriculars
other activities
student government
student newspaper or other
debate
fraternity
no student activities

IV. MORAL: having to do with moral character or moral commitment.

Composed of three meso categories: (1) moral character; (2) work ethic; and (3) social consciousness.

A. Moral character

responsible
good conduct
of the highest/excellent character

good/sound character
moral integrity
integrity
dignity/strength of character
sound/balanced judgement
has ideals/convictions
courageous
intellectual/academic integrity
honest
intellectual honesty
not intellectually honest
spiritual or religious
search for meaning
earnest
loyal/dedicated
human/humanity
unselfish
oustanding human being

B. Work ethic

hard worker/industrious/diligent
disciplined/self-discipline
ambitious/driven/determined
dependable/conscientious/reliable
doesn't push himself (lazy?)
not lazy

C. Social consciousness

sense of social responsibiity
ethical concerns
links work to social world
social consciousness
student political organizations
not involved in campus politics

V. PERSONAL: having to do with personality or psyche

Composed of three meso categories: (1) personality; (2)interpersonal relations; and (3) psychological and emotional well-being.

A. Personality

can laugh at himself
sense of humor
no sense of humor
refreshing/fresh spirit
dynamic/alive/energy
excited/eager/enthusiastic
not aggressive
witty
interesting (person and work)
provocative/challenging/inspiring
intense/direct
serious
patient
impatient
strong personality
sensitive
open-minded
forceful personality
relaxed/calm
quiet/shy
positive/constructive/realistic
rather boring
stubborn
self-reliant/independent
unusual
naive
adaptable
introspective/thoughtful/reflective
aggressive

B. Interpersonal relations

cooperative/teachable
knows when he's being pompous
gets along well with others
pleasant/friendly
irritating
thoughtful/considerate
sincere
arrogant/pompous
not arrogant
takes himself too seriously at times

C. Psychological and emotional well-being

self-assured
balanced personality
well-adjusted
immature
mature
knows himself, his capabilities/limits
insecure

Appendix B. Sample letter of recommendation, 1965

Mr. X did consistent A work for me for two semesters. He is a lucid thinker, articulate in discussion, writes precisely and maturely. There is something very engaging about his mind and personality. He is lively, witty, and at the same time softspoken, genuinely unassuming. He is used to thinking vigorously on his own, and yet he is very much open to learning from others. He has shown, moreover, both in his work as a student and as editor of the [university's literary magazine], the kind of serious engagement in literature that a future professional should have. I would expect him to excel in any graduate program and to be a very competent scholar-critic and teacher afterward.

References

Abbott, Andrew, 1988. The System of Professions: An Essay on the Division of Expert Labor. University of Chicago Press, Chicago, IL.
American Council on Education, 1970. College and University Faculty: A Statistical Description. A.C.E. Research Report 5. Washington, DC.
Arrow, Kenneth, Bowles, Samuel, Durlauf, Steven (Eds.), 1999. Meritocracy and Economic Inequality. Princeton University Press, Princeton, NJ.
Babbie, Earl, 1989. The Practice of Social Research, fifth ed. Wadsworth Publishing Co, New York, NY.
Becher, Tony, 1989. Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines. The Society for Research into Higher Education and Open University Press, Milton Keynes, UK.
Bender, Thomas, Schorske, Carl (Eds.), 1998. American Academic Culture in Transformation: Fifty Years, Four Disciplines. Princeton University Press, Princeton, NJ.
Bender, Thomas, 1997. Politics, intellect, and the American university, 1945–1995. Daedalus 126, 1–38.
Berelson, Bernard, 1960. Graduate Education in the United States. McGraw-Hill Book Company, New York, NY.
Bourdieu, Pierre, 1996. The State Nobility: Elite Schools in the Field of Power. Stanford University Press, Stanford, CA.
Bourdieu, Pierre, 1988. Homo Academicus. Polity Press, Cambridge, UK.
Brim, Orville, Glass, David, Neulinger, John, Firestone, Ira, 1969. American Beliefs and Attitudes about Intelligence. Russell Sage Foundation, New York, NY.
Collins, Randall, 1979. The Credential Society: A Historical Sociology of Education and Stratification. Academic Press.

DeVane, William, 1965. Higher Education in Twentieth-century America. Harvard University Press, Cambridge, MA.
Dobbin, Frank, Sutton, John, Meyer, John W., Scott, Richard, 1993. Equal opportunity law and the construction of internal labor markets. American Journal of Sociology 99, 396–427.
Feller, William, 1957. An Introduction to Probability Theory and Its Applications, Vol. 1. John Wiley and Sons, New York, NY.
Fischer, Claude, Hout, Micheal, Sanchez-Jankowski, Martin, Swidler, Ann, Lucas, Samuel, 1996. Inequality by Design: Cracking the Bell-Curve Myth. Princeton University Press, Princeton, NJ.
Friedland, Roger, Alford, Robert, 1991. Bringing society back in: symbols, practices, and institutional contradictions. In: DiMaggio, P.J., Powell, W.W. (Eds.), The New Institutionalism in Organizational Analysis. University of Chicago Press, Chicago, IL, pp. 232–263.
Geiger, Roger, 1993. Research and Relevant Knowledge: American Research Universities Since World War II. Oxford University Press, Oxford, UK.
George, Gerald, 1967. Bearers of the Torch. Unpublished manuscript, Woodrow Wilson Foundation, Princeton, NJ.
Gieryn, Thomas, 1999. Cultural Boundaries of Science: Credibility on the Line. University of Chicago Press, Chicago, IL.
Harris, Seymour, 1972. A Statistical Portrait of Higher Education. McGraw-Hill Book Company, New York, NY.
Hofstadter, Richard, 1961. American Higher Education: A Documentary History, Vol. 1 and 2. University of Chicago Press, Chicago, IL.
Jencks, Christopher, Riesman, David, 1968. The Academic Revolution. University of Chicago Press, Chicago, IL.
Jencks, Christopher, Phillips, Meredith (Eds.), 1998. The Black-White Test Score Gap. Brookings Institution Press, Washington, DC.
Karen, David, 1985. Who Gets Into Harvard? Selection and Exclusion at an Elite College. PhD dissertation, Department of Sociology, Harvard University.
Kish, Leslie, 1995. Survey Sampling. John Wiley and Sons, New York, NY.
Ladd, Everett, Lipset, Seymour, 1975. The Divided Academy: Professors and Politics. The McGraw-Hill Book Company, New York, NY.
Lamont, Michèle, 2000. What Stands for Quality: Criteria of Evaluation of Knowledge in the Social Sciences and the Humanities. Unpublished manuscript, Department of Sociology, Princeton University.
Lamont, Michèle, Lareau, Annette, 1988. Cultural capital: allusions, gaps and glissandos in recent theoretical developments. Sociological Theory 6, 153–168.
Lamont, Michèle, Kaufman, Jason, Moody, Michael, 2000. The best of the brightest: definitions of the ideal self among prize-winning students. Sociological Forum 15, 187–224.
Larson, Magali Sarfatti, 1977. The Rise of Professionalism: A Sociological Analysis. University of California Press, Berkeley, CA.
Lazarsfeld, Paul, Thielens, Wagner, 1958. The Academic Mind: Social Scientists in a Time of Crisis. Free Press, New York, NY.
Lemann, Nicholas, 1999. The Big Best: The Secret History of the American Meritocracy. Farrar, Straus and Giroux, New York, NY.
Lewis, Lionel, 1998. Scaling the Ivory Tower: Merit and Its Limits in Academic Careers. Transaction Publishers, New Brunswick, NJ.
Merton, Robert, 1949. Social Theory and Social Structure. Free Press, New York, NY.
Oleson, Alexandra, Voss, John (Eds.), 1979. The Organization of Knowledge in Modern America, 1860–1920. Johns Hopkins University Press, Baltimore, MD.
Parsons, Talcott, 1951. The Social System. Free Press, New York.
Rosenhaupt, Hans, 1969. University: a Princeton Quarterly 40, 17–22.
Ross, Dorothy, 1990. The Origins of American Social Science. Cambridge University Press, New York, NY.
Sen, Amartya, 1999. Merit and justice. In: Arrow, K., Bowles, S., Durlauf, S. (Eds.), Meritocracy and Economic Inequality. Princeton University Press, Princeton, NJ, pp. 5–16.

Schudson, Michael, 1972. Organizing the 'meritocracy': a history of the college entrance examination board. Harvard Educational Review 42, 34–69.
Snow, Charles Percy, 1959. The Two Cultures and the Scientific Revolution. Cambridge University Press, New York, NY.
Synnott, Marcia, 1979. The Half-opened Door: Discrimination and Admissions at Harvard, Yale, and Princeton, 1900–1970. Greenwood Press, Westport, CT.
Tittle, Charles R., Rotolo, Thomas, 2000. IQ and stratification: an empirical evaluation of Herrnstein and Murray's social change argument. Social Forces 79, 1–28.
Veysey, Laurence, 1965. Emergence of the American University. University of Chicago Press, Chicago, IL.
Wechsler, Harold, 1977. The Qualified Student: A History of Selective College Admission in America. John Wiley and Sons, New York, NY.
Whitley, Richard, 1984. The Intellectual and Social Organization of the Sciences. Oxford University Press, Oxford, UK.
Wilson, Logan, 1979. American Academics Then and Now. Oxford University Press, Oxford, UK.
Young, Michael, 1958. The Rise of Meritocracy. Thames and Hudson, London.

Angela Tsay is on leave from her graduate studies and has spent the last three years in COO and other operations positions at several successful startups. She currently resides in the Bay Area.

Michèle Lamont is Professor of Sociology at Harvard University, having recently moved from Princeton University. She is the author of *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration* (2001 C. Wright Mills Prize; 2001 Mattei Dogan Prize). A former Chair of the Culture Section of the American Sociological Association, she is now serving as the Chair-Elect of the Theory Section and as a member of the Executive Committee of the Council for European Studies. She is currently a Fellow at the Center for Advanced Study in the Behavioral Sciences, working on a book on definitions of excellence used by members of funding panels in the social sciences and the humanities in the United States (funded by NSF). A former Guggenheim fellow, she is a coeditor of the series *Princeton Studies in Cultural Sociology*, published by Princeton University Press.

Andrew Abbott is Gustavus F. and Ann M. Swift Distinguished Service Professor at the University of Chicago. He has written on professions and work, on education and disciplines, and on theory and methodology. He is currently writing a theoretical book on time and social structure.

Joshua Guetzkow is a PhD candidate at Princeton University. His research interests include political criminology, the sociology of knowledge and education. His dissertation research examines the impact of framing on historical changes in sentencing and social welfare policies in the U.S.