Balance between merit and equity in academic hiring decisions

Judgemental content analysis applied to the phraseology of Australian tenure-stream advertisements in comparison with Canadian advertisements

Gregory J Boyle
Bond University, Queensland

John J Furedy
University of Toronto, Canada

David L Neumann, H Rae Westbury & Magnus Reiestad
Griffith University, Queensland

The wording of university academic job advertisements can reflect a commitment to equity (affirmative action) as opposed to academic merit in hiring decisions. The method of judgemental content analysis was applied by having three judges rate 810 Australian tenure-stream advertisements on seven-point magnitude scales of equity and merit. The influence of time (Years: 1970-1973; 1984-1987; 2000-2003), institution (major research universities, the self-designated Group of Eight (Go8); colleges of advanced education and institutes of technology; regional and distance education institutions), as well as academic discipline (physical sciences and technology; social sciences; humanities) on ratings were also examined. Inter-rater reliabilities were high (≥ 0.92), and the 'equivalence hypothesis' (that merit and equity are the same) was not supported. Merit and equity criteria increased over time and were influenced by institution type and academic discipline, although in different ways. While some effects could be viewed as being due to rational policy decisions, other significant effects suggested influences that are more difficult to explain. University administrators need to be sensitive to the balance between merit and equity when formulating hiring policies.

Introduction

In the early 1970s, the introduction of public advertisements for all academic positions in North American and Australian universities brought about a substantial change in the hiring process. In principle, the reform increased the possibility that academic merit (i.e. what the candidate knew rather than who the candidate knew) would provide the basis of selection for academic positions. In the early 1980s, in addition to the goal of merit in appointments, a goal variously described as 'diversity', 'affirmative action', or 'equity' was adopted to a greater or lesser extent by universities. Affirmative action policies applied to tenure-stream academic appointments involve a certain degree of preference for individuals who are members of particular sub-

groups (e.g. based on gender or ethnic background). These policies were intended to produce 'equity' in outcomes of the proportions of competition winners from within the specified subgroups, rather than just in opportunities to apply for such competitions.

Any preferential treatment given to certain applicants based on their subgroup membership can be interpreted as being incompatible with a merit-based system (Kranz 1974). The conflict between group-preference and merit policies is especially problematic when merit itself involves complex, expert judgements, as is the case for academic tenure-stream positions. The degree of commitment to employment equity policy in tenure-stream hiring can vary considerably along the equity-merit continuum. In some universities, for example, 'equity officers' exert direct pressure on departmental chairs regarding both the wording of job advertisements and decisions on appointments.

Institutional commitment to affirmative action is partly influenced by conscious decisions made by university administrations. However, there may be less conscious, latent influences that play a role in determining the point of 'balance' universities settle on regarding the somewhat conflicting principles of equity and merit. It is possible to study systematically the differential adaptations of universities (and even within university sectors, such as the physical sciences and technology versus the humanities). This approach involves a content analysis of the phraseology of tenure-stream advertisements to assess differences in the relative degree of emphasis placed on merit and equity respectively, as factors in hiring within different institutions of higher learning.

Furedy *et al.* (2001) assessed the relative emphasis of merit and equity in 519 advertisements for tenure-stream positions at Canadian universities. They reported that the inclusion of merit criteria increased across the disciplines in order of the humanities, social sciences, and physical sciences (which they labelled as "hard sciences"), respectively. Merit requirements also increased over time and according to the university 'mission' of undergraduate, comprehensive, and medical/doctoral, respectively. However, several interactions also emerged to suggest that merit criteria in advertisements were influenced differently according to the combination of discipline, university mission, and university location. As with the merit findings, the inclusion of equity criteria in the advertisements was influenced by university mission in that it increased across the undergraduate, comprehensive, and medical/doctoral institutions, respectively, these being categories used to classify uni-

versities in Canada. Also similarly, interactions emerged between the various factors of time, location, mission, and discipline. However, the number of interactions seemed to be far more extensive for equity than they were for merit. On the basis of these findings, the authors suggested that many latent influences exert their influence on merit and particularly equity information in university tenure-stream advertisements.

As shown in the study by Furedy *et al.* (2001), the relationship between merit and equity can be examined by considering the patterns of outcomes of the merit and equity measures. On the one hand, the various variables such as time and discipline may influence merit and equity in the same way and produce a similar pattern of results for each (the 'equivalent hypothesis'). On the other hand, a different pattern of results would indicate that the external variables do not influence merit and equity in the same way. In their analysis of Canadian tenure-stream advertisements, Furedy *et al.* (2001) reported that the pattern of results did differ between merit and equity, and quite dramatically at times. The investigators concluded that merit and equity can be influenced by conscious and unconscious (latent) pressures in different ways. Further, these influences can ultimately influence the hiring decisions for academic positions (see also discussion by Furedy & Furedy 2003).

The extent to which the findings of Furedy *et al.* (2001) apply to tenure-stream advertisements at Australian universities is not known. The present study thus aimed to examine patterns in the inclusion of merit and equity criteria in Australian university advertisements across different time periods, disciplines, and types of institutions. The examination of the results for merit and equity in isolation will indicate the extent to which the inclusion of this information is influenced by these factors. Moreover, merit and equity criteria may be directly compared. The equivalence hypothesis implies that, aside from errors of measurement, merit and equity measures should be strongly correlated. Moreover, the equivalence hypothesis would lead to the prediction that merit and equity will be influenced in the same way by the variables of Time, Discipline, and Institution and their interactions.

The method of judgemental content analysis of tenure-stream advertisements

Selection and categorisation

A total of 810 tenure-stream advertisements were sourced from archives of the Wednesday and Saturday

editions of *The Australian* newspaper. Only advertisements that were made during one of three time periods were sourced. The first time period was 1970 to 1973, when the use of open advertisements for all tenure-stream positions was instituted. The second time period of 1984 to 1987 is a pre-amalgamation period because it spans the time just before the so-called Dawkins reforms. After this time period, colleges of advanced education and institutes of technology were either amalgamated with universities (e.g. Melbourne College of Advanced Education became part of the University of Melbourne) or were elevated to full university status (e.g. Queensland Institute of Technology became the Queensland University of Technology). The third time period was 2000 to 2003 and it can be considered to be a post-amalgamation period in Australia. The three time periods thus span the levels of open advertisements, pre-amalgamation, and post-amalgamation.

The tenure-stream advertisements within each time period were also classified according to discipline and the academic status of the institution. The disciplines were classified as physical sciences and technology, the social sciences, or the humanities, respectively, in terms of their quantitative scientific demands. The institutional type was classified as Group of Eight (Go8) universities, combined Institutes of Technology and Colleges of Advanced Education (CAET), or Regional and Distance Education tertiary institutions (RDE), respectively. However, there were some slight exceptions with the institution categorisation scheme. For example, Monash (Go8) has both a regional campus, formally a stand-alone CAE, and distance education. The institution categorisation was essentially one of academic status or institutional prestige.

Ratings of equity and merit

The 810 tenure-stream advertisements were rated by three independent judges according to the respective levels of merit and equity criteria they contained. The relative emphases on merit and equity in tenure-stream advertisements cannot be assessed by means of completely objective methods such as counting the frequency of certain words. The content analysis that has to be applied (see Furedy *et al.* 2001 for details) must be more subjective or 'judgemental', and essentially involves the psychophysical method of magnitude estimation by informed observers. The content analysis involved a degree of subjectivity and reliance on rater judgement.

The methods of content analysis utilised by Ilic (1999) were improved upon by (a) having three rather

than one judge, (b) undertaking extensive discussions of rating methods to ensure adequate reliability with the ratings, (c) using a quasi-random order of ratings with respect to the categories of time, discipline, and institution to avoid sequence effects, and (d) removing university and department names from the job advertisements so that they would not bias the ratings made. Ratings were made based on the identification of 'markers' associated with equity and merit. Merit markers included: requests for published reprints and preprints; demonstrated research ability; graduate level teaching experience; strong academic background, and evidence of ability to develop a viable, externally-funded research program. Markers associated with high equity wording included: candidates from specified subgroups being given 'preference' or 'especially' being encouraged to apply. The judges rated the wording of the advertisements using separate seven-point scales to score the degree to which the advertisements emphasised merit and equity. Higher ratings indicate a higher endorsement of merit or equity in the advertisement.

Training was used to enhance reliability of the ratings. For the training, a further 90 tenure-stream advertisements were sourced from archives of the Wednesday and Saturday editions of The Australian newspaper, ensuring that the date of the advertisement did not overlap with those examined in this study. The three raters independently made ratings of equity and merit for each advertisement. Next, the ratings made by each rater were compared. When discrepancies of two or more points on the seven-point scale occurred, the possible reasons for the differences were discussed. As a result, the individual raters were able to be more comparable in their interpretation of relative strength of equity and merit markers on the measurement scale.

Evaluation of the reliability of the ratings

The reliability between the raters was assessed by examining the observed inter-rater reliability coefficients. The mean inter-rater reliability estimates of 0.96 and 0.93 for merit and equity ratings, respectively, indicated a high reliability in the ratings made. A further check was conducted by calculating the correlation in ratings between each pair of raters. The individual correlations for merit ratings between raters were 0.89 (A with B), 0.92 (A with C) and 0.89 (B with C). Likewise, for equity ratings, correlations were 0.92 (A with B), 0.86 (A with C) and 0.82 (B with C). Although the method of judgemental content analysis is not strictly objective, the use of trained raters thus yielded high

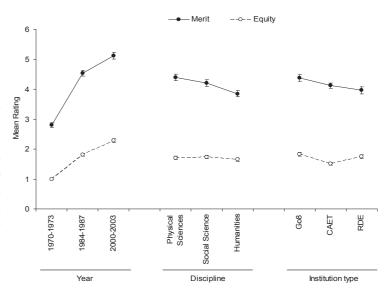inter-rater reliability estimates for both merit and equity.

Results

Relationship between merit and equity

Before statistical analyses, the merit and equity ratings were each averaged across all three raters due to the established consistency among them. The relationship between merit and equity ratings was initially examined by looking at the correlation between the two measures. For the entire set of ratings, there was a moderate, though statistically significant positive association between merit and equity ($r = .35, p < .001$). The association reflected that higher merit ratings were associated with higher equity ratings.

To compare merit and equity directly, the statistical approach called a factorial analysis of variance (ANOVA) was used. This approach included merit and equity ratings as correlated-samples Measures factor in the analysis (cf. Hair *et al.* 2006). The ANOVA also included the between-samples variables of Time (1970-1973, 1984-1987, 2000-2003), Discipline (physical sciences, social sciences, humanities), and Institution (Go8, CAET and RDE). The ANOVA can examine the effect that each variable has on its own (called the main effects) as well the combination of variables (called the interactions) in terms of their association with merit and equity ratings. Main effects and interactions emerge if the test statistic associated with them has a probability p < .05.

The resulting ANOVA employed a $2 \times 3 \times 3 \times 3$ (Measures × Time × Discipline × Institution) design. The equivalence hypothesis leads to the prediction that the Time, Discipline, and Institution factors will influence merit and equity in the same way. Thus, there should be no interaction between the Measures factor and these other factors. However, the analysis yielded a statistically significant Measures × Time × Discipline × Institution interaction, $F(8,1566) = 3.37, p = .001$. A significant Measures × Time × Institution interaction was also found, $F(4,1566) = 3.56, p = .007$. Finally, the Measures factor interacted significantly with Time, $F(2,1566) = 33.13, p < .001$, Discipline, $F(2,1566) = 6.22, p < .002$, and Institution, $F(2,1566) = 4.57, p = .011$. These latter two-way interactions are represented

Figure 1. Mean ratings for merit and equity when the advertisements were grouped into either Time (Year), Discipline, and Institution. Error bars indicate the standard error of the mean.
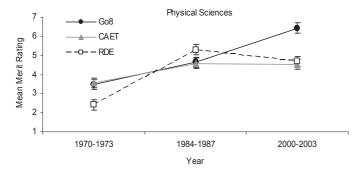

in Figure 1 for which it can be seen that the pattern of results across the levels of each variable show some similarities, but also differences for merit and equity. An example of one difference is that merit ratings declined in order of Go8, CAET, and RDE institutions. In contrast, equity ratings declined in order of Go8, RDE, and CAET institutions.
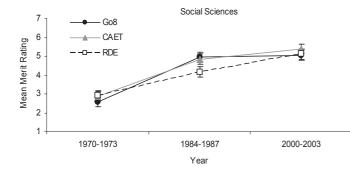
The interactions involving the Measures factor show that, contrary to the equivalence hypothesis, merit and equity were influenced differently by the various variables of Time, Discipline, and Institution. The exact way in which they differed can be determined by comparing the pattern of results in how these variables influence merit and equity in isolation. The subsequent statistical analyses were conducted with this aim. Separate ANOVAs were run on merit ratings and equity ratings and each employed a $3 \times 3 \times 3$ (Time × Discipline × Institution) between-samples ANOVA design.
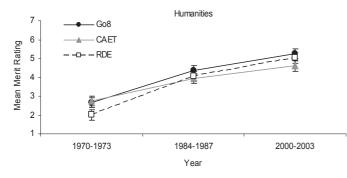
Merit

The merit ratings averaged across each of the levels for the variables of Time, Discipline, and Institution are shown in Figure 1. The statistical analyses showed a main effect for Time, $F(2\ 783) = 187.01, p < .001$, indicating that ratings varied across the three time periods. To work out exactly which time periods differed statistically, the t-test was used with a Bonferroni adjustment applied to correct for possible statistical error associated with multiple comparisons. These analyses

Figure 2. Mean merit ratings grouped by discipline for each institution type as a function of the year of the advertisement. Error bars represent the standard error of the mean.

indicated that all years differed significantly from each other, all $ts > 4.17, p < .001$. Thus, merit information increased from the open advertisement years (1970-1973), to the pre-amalgamation years (1984-1987), and again to the post-amalgamation years (2000-2003).

The disciplines also differed in merit ratings, as shown by a main effect for Discipline, $F(2,783) = 10.02, p < .001$. Figure 1 indicates that merit information in advertisements declined across the disciplines in order of the physical sciences, social sciences, and humanities. However, only the difference between the ratings for the physical sciences and humanities was statistically significant, $t(538) = 3.59, p < .001$.

The final main effect for merit ratings concerned the academic status of the institution. The main effect was significant, $F(2,783) = 5.62, p < .01$. It was due to a statistically significant higher merit rating for the Go8 universities than for the RDE universities, $t(538) = 2.51, p = .012$.

Although each of the variables of time, discipline, and institution exerted an influence on merit ratings on their own, they also interacted in their effects. This was shown by a Time × Institution interaction, $F(4,783) = 3.14, p < .05$, and a three-way Time × Discipline × Institution interaction, $F(8,783) = 4.17, p < .001$. The interpretation of three-way interactions can be complex. It is eased by comparing the effects of two variables when the third is kept constant. For the present purposes, it makes sense to compare the different levels of institution at each year for each individual discipline. Figure 2 shows the merit ratings grouped by discipline. As can be seen, the social sciences and humanities disciplines showed a different pattern to the physical sciences.

Further statistical analyses were conducted to examine the three-way interaction. Separate 3 × 3 (Time × Institution) ANOVAs were run for each discipline. As expected, a Time × Institution interaction emerged for the physical sciences discipline, $F(4,261) = 8.43, p < .001$, and not for the other disciplines, all $Fs < 1.51, p > .05$. For the physical sciences, merit ratings increased for all institutions from 1970-1973 to 1984-1987, all $ts > 2.89, p < .005$. From 1984-1987 to 2000-2003, merit ratings increased for the Go8 universities, $t(58) = 5.25, p < .001$, but not for the CAET and RDE institutions, both $ts < 1.31, p > .05$. To summarise this pattern of results, merit ratings increased across each time period for all institutions in the social sciences and humanities and for the Go8 institutions in the physical sciences. However, there was no increase in merit ratings from the pre-amalgamation years (1984-1987) to the post-amalgamation years (2000-2003) for the CAET and RDE institutions in the physical sciences discipline.

Equity

The equity ratings for each individual variable of Time, Discipline, and Institution are shown in the lower portion of Figure 1. As can be seen, the values for the equity ratings are lower than for the merit ratings and they showed some similarities, but also some differences in the patterns across each of the variables. The analyses showed a main effect of Time, $F(2,783) = $

209.03, $p < .001$. Similar to the merit ratings, there was a significant increase in equity ratings from 1970-1973 to 1984-1987, $t(538) = 18.49$, $p < .001$, and again from 1984-1987 to 2000-2003, $t(538) = 5.82$, $p < .001$. Unlike the merit ratings, there was no significant main effect of Discipline for equity ratings. A main effect of Institution was found, $F(2,783) = 13.30$, $p < .001$, and it indicated that equity ratings differed among the three groups of institutions. Equity ratings were lower for the CAET institution than the Go8 and RDE institutions, both $ts > 3.07$, $p < .002$. Equity ratings for the Go8 and RDE institutions did not differ, $t(538) = .91$, $p > .05$.
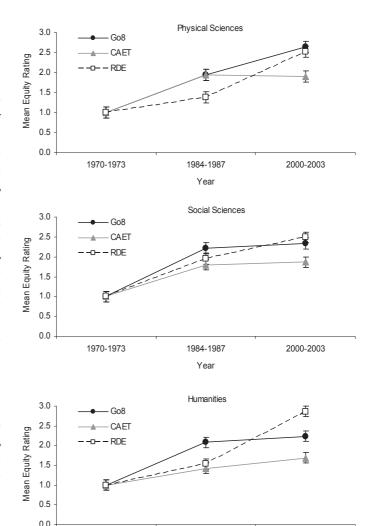
An examination of the interactions among the variables revealed that only the Time × Institution interaction was statistically significant, $F(4,783) = 12.64$, $p < .001$. To facilitate comparisons with the merit ratings, Figure 3 shows the ratings for time and institution grouped by discipline type. However, further analyses on the interaction were averaged across the disciplines because, unlike the merit ratings, there was no significant three-way interaction. As can be seen in Figure 3, all institutions increased in equity ratings from 1970-1973 to 1984-1987, all $ts > 9.12$, $p < .001$. From 1984-1987 to 2000-2003 the increase was significant for RDE, $t(178) = 6.90$, $p < .001$, but did not reach statistical significance for the Go8 or CAET institutions, both $ts > 2.17$, $p > .03$.

Discussion

Summary and interpretation of findings

Taken together, the results show that the nature of merit and equity information in advertisements for tenure-stream academic positions differed across time, institution, and discipline. The present study yielded a significant, albeit moderate correlation between merit and equity ratings to indicate that higher merit criteria in advertisements were associated with more equity information. In contrast, a previous Canadian study yielded a non-significant correlation (Furedy *et al.* 2001). The present findings highlight a difference between Australian and North American advertisements suggesting that in Australia the conflict between implementing merit and equity principles has not been as marked as in the Canadian context. The simi-

Figure 3. Mean equity ratings grouped by discipline for each institution type as a function of the year of the advertisement. Error bars represent the standard error of the mean.

larity between merit and equity in some of the findings, such as the increase in ratings across the three time periods for both, also supports this conclusion. However, the equivalence hypothesis that merit and equity are the same did not receive complete support because there were several differences in the patterns of results between the merit and equity ratings. For this reason, the effects of time, institution, and discipline were examined separately for merit and equity.

As to the merit and equity results examined individually, some results would appear to reflect predictable adapations by the universities. Examples of such results include an increase of merit requirements over time (e.g. as the proportion of candidates having post-doctoral experience has increased), discipline (e.g. with

the physical sciences requiring more extensive academic and research qualifications than the humanities in say post-doctoral experience), and institution (e.g. in more prestigious institutions requiring higher academic merit). In the case of equity, the effect of time (increased equity emphasis over time) can be readily justified. It would appear to reflect that academic human resource administrators were becoming more aware of the importance of proactively emphasising equity in the merit/equity balance, rather than just being content to eliminate non-merit-based sources of bias, such as prejudice against female academic applicants.

Interaction effects involving the three variables of time, institution, and discipline with merit and equity were observed. An example was the significant interaction between time and institution. For equity, the CAET and Go8 institutions fell behind the RDE institutions in their equity emphasis in the later years. This suggests that equity policies in academic appointments have been further developed in regional institutions, whereas they have not substantially increased for the Go8 or CAET institutions. It would be interesting to determine whether this has had any differential impact in actual hiring decisions at the different types of institutions in the later years. An interaction between time and institution was also observed for merit. However, it may be unlikely that this interaction reflects the same influences to that of equity. For merit, the interaction between time and institution was limited to the physical sciences discipline and was contrary to the pattern found with equity. In the last time period, the CAET and RDE institutions fell substantially behind the Go8 institutions for merit criteria. Indeed, the merit ratings appeared to decline somewhat for the CAET institutions in this time period.

The present research had some limitations that could be addressed in future research. The three categories of physical sciences, humanities, and social sciences captured many of the disciplines at university, but could be expanded in future research to include other disciplines, such as the creative and performing arts. The present study also used three raters and obtained a high degree of consistency among them in the ratings made. However, future research could include more raters to provide an even better check on the reliability and appropriateness of the ratings for equity and merit. Finally, future research could apply the methods used here to examine more recent advertisements so that, when combined with the present findings, the examination of the balance between merit and equity in tenure-stream advertisements can be extended over a longer time frame.

Conclusion

It appears that in adapting to the conflicting requirements of 'balancing' merit and equity considerations, Australian universities have so far been able to adopt more rationally justifiable policies than those that appear to be operational in Canadian universities. Nevertheless, in terms of the practical implications of the present findings, university human resource administrators responsible for hiring policies in relation to academic staff need to consider potentially discriminative influences such as affirmative action that influence the wording of tenure-stream advertisements, over and above those influences that do not discriminate in favour of candidates from particular subgroups. Further research into university hiring policies would seem warranted, especially in regard to its impact on student performance and academic outcomes.

Gregory Boyle is Professor of Psychology at Bond University, Queensland, Australia.

John Furedy is Emeritus Professor of Psychology at University of Toronto, Canada.

David Neumann is Senior Lecturer in Psychology at Griffith University, Queensland, Australia.

Rae Westbury and Magnus Reiestad are doctoral students at Griffith University, Queensland, Australia.

Acknowledgement

References

Furedy, JJ, Fidler, S, Morgenstern, Y & Tryhor, W 2001, 'Judgmental content analysis of Canadian tenure-stream advertisements to assess latent influences on institutional commitment to affirmative action', *Technical Report #1* for the Donner Canadian Foundation. Available from: <http://www.psych.utoronto.ca/~furedy/Papers/me/JUDGM6.doc>.

Furedy, JJ & Furedy, CP 2003, 'Tenure-stream advertisements before and after the 1995 'Common Sense Revolution' in Ontario', *Society for Academic Freedom and Scholarship Newsletter,* 33, pp. 4-9.

Hair, JF, Black, WC, Babin, BJ, Anderson, RE & Tatham, RL 2006, *Multivariate data analysis* (6th ed.), Upper Saddle River, NJ, Pearson/Prentice-Hall.

Ilic, K 1999, 'Merit and equity in Canadian tenure-stream position advertisements: Effects of time, location, mission, and discipline hardness', Unpublished undergraduate thesis, Department of Psychology, University of Toronto.

Kranz, H 1974, 'Are merit and equity compatible?', *Public Administration Review,* 34, pp. 434-440.