Features:

*Students of Color and Doctoral Programs: Factors Influencing the Application Decision in Higher Education Administration*
— Michael C. Poock

*Employing Quantitative Models of a Qualitative Admissions Process: Uncovering Hidden Rules, Saving Time, and Reducing Bias*
— Philip M. Sadler and James K. Hammerman

*A Study of Canadian and American Admissions Equity at Brigham Young University*
— Jeffery M. Tanner and R. Wayne Shute

*Educational Consortia — A Longitudinal Study*
— Marybelle C. Keim

*Level of Math Preparation in High School and Its Impact on Remedial Placement at an Urban State College*
— Jeff E. Hoyt

The Forum:

International Q&A, Commentary, Book Reviews, and Letters to the Editor
When quality counts, don’t take chances – take TOEFL®.

The TOEFL test has helped more than 2,400 colleges and universities select qualified students since 1965. Over 11 million students have taken the exam in 170 countries around the world. Backed by years of research and experience, no other test is as reliable a standard for measuring a nonnative speaker’s ability to use English at the college level.

Now on computer, the TOEFL test offers faster test results and features important innovations – including a multimedia Listening section and a compulsory written essay – designed to give institutions a more complete picture of each candidate’s abilities.

Why settle for second best? With so much at stake in each admissions and placement decision, insist on TOEFL, the international standard for 35 years.

www.toefl.org

TOEFL means quality in any language.
Features

Students of Color and Doctoral Programs: Factors Influencing the Application Decision in Higher Education Administration .............................................. 2

Michael C. Poock looks at the factors affecting students of color’s decisions to apply to certain doctoral programs in higher education administration.

Employing Quantitative Models of a Qualitative Admissions Process: Uncovering Hidden Rules, Saving Time, and Reducing Bias ........................................ 8

This study performed by Philip M. Sadler and James K. Hammerman quantitatively models the admissions process at Harvard University’s Graduate School of Education’s Doctoral Program in Learning and Teaching.

A Study of Canadian and American Admissions Equity at Brigham Young University .................................. 25

Jeffery M. Tanner and R. Wayne Shute determine whether admissions standards differ for American and Canadian students at Brigham Young University.

Educational Consortia — A Longitudinal Study ........................................... 30

Marybel C. Keim examines consortia in higher education.

Level of Math Preparation in High School and Its Impact on Remedial Placement at an Urban State College ........................................ 37

Jeff E. Hoyt’s study focuses on remedial placement and the math preparation of recent high school graduates. There were concerns about the rigor of high school math programs and the ability of students to retain math skills.

The Forum

Commentary
Why We Can’t Break Out of the Box ......................................................... 44

International Q&A .................................................................................. 46

Book Reviews
The Shape of the River .................................................................................. 49
Taking Women Seriously: Lessons and Legacies for Educating the Majority ........................................ 50
Accommodations in Higher Education under the Americans with Disabilities Act (ADA) ........ 51

Letters to the Editor .................................................................................... 52

Founded in 1910, AACRAO is a nonprofit, voluntary, professional education association of degree-granting postsecondary institutions, government agencies, higher education coordinating boards, private educational organizations, and education-oriented businesses. The mission of the Association is to provide leadership in policy initiation, interpretation, and implementation in the global educational community. This is accomplished through the identification and promotion of standards and best practices in enrollment management, instructional management, information technology, and student services.

• AACRAO does not endorse the products or services of any advertiser. The advertiser assumes responsibility and liability for the content of any advertising. The opinions expressed by the authors do not necessarily reflect the views of AACRAO.

• AACRAO adheres to the principles of nondiscrimination and equality without regard to race, color, creed, gender, sexual orientation, age, religion, disability, or national origin.

• Questions on subscriptions, back issues, advertising, and membership information should be addressed to AACRAO, One Dupont Circle, NW, Suite 520, Washington, DC 20036-1135; phone (202) 293-9161; fax (202) 872-8857. Printed in the USA. Periodical postage is paid in Washington, DC.

• College & University is available on microfilm though UMI/Periodicals, P.O. Box 1346, Ann Arbor, MI 48106-1346; Phone: (734) 761-4700.

POSTMASTER: Send address changes to College & University, One Dupont Circle, NW, Suite 520, Washington, DC 20036-1135. Copyright © 1999 American Association of Collegiate Registrars and Admissions Officers.
The C&U Editorial Board welcomes manuscripts for publication in College & University, AACRAO's scholarly research journal. AACRAO members are especially encouraged to submit articles pertaining to their own experiences with emerging issues or innovative practices within the profession.

The Board also welcomes comments on articles, timely issues in higher education, and other topics of interest to this journal's readers in the form of letters to the editor or longer guest commentary. We especially invite AACRAO members to participate in reviewing books.

Manuscript Preparation

Manuscripts for feature articles should be no longer than 4,500 words. Manuscripts for guest commentary and book reviews should not exceed 2,000 words. Letters to the editor will ordinarily be limited to 200 words.

All submissions must be saved to an IBM-compatible disk (Microsoft Word, preferably) and include a hard-copy original printed on 8.5" x 11" white paper. Because the Board has a blind review policy, the author's name should not appear on any text page. A cover sheet should include the title of the manuscript, author's name, address, phone and fax number, and e-mail address.

References should be formatted in the author-date style and follow guidelines provided on page 526 of The Chicago Manual of Style, 14th edition. A list of references should appear at the end of the article. Text citations also follow the author-date format; examples may be found on page 641 of the Manual. For more information or for samples, please contact the C&U Editor.

Essential tables and charts should be included on separate pages at the end of the manuscript. All graphics should be submitted on clean, reproducible, or camera-ready paper.

All submissions are accepted for publication with the understanding that the College & University editors reserve the right to edit for clarity and style. Please do not submit articles that are under consideration for publication by another periodical.

Authors whose manuscripts are selected for publication will be asked to submit a short biographical statement and an abstract of their article, each no more than thirty-five words.

Submit manuscripts, letters, and direct inquiries to—

Roman S. Gawkoski
C&U Editor
Office of the Registrar
Marquette University
PO Box 1881
Milwaukee, WI 53201-1881

phone: (414) 782-3855; fax: (414) 288-1773; e-mail: 9761gawkoski@vms.csd.mu.edu

Editorial Procedure

The editor will acknowledge receipt of manuscripts (letters will not be acknowledged) and will forward them to members of the C&U Editorial Board for review. The Board will consider the appropriateness of the article for AACRAO's membership, the current needs of the professions, the usefulness of the information, the nature and logic of the research methodology, clarity, and the style of presentation.

This review may take as long as three months, after which the C&U editor will inform the author of the manuscript's acceptance or rejection.
**Editor's Note**

**THESE HAPPEN EVERY SPRING —**

- The ice and snow on our waterways and fields have melted
- The vernal equinox is past and we're headed toward the longest day
- Trees and grass are greening
- Easter is a late memory and Christmas shopping has begun
- Crocus and Jonquil are abloom
- The 1999 AACRAO annual conference is over, highly successful with numerous sessions of interest to the membership

**BUT THIS DOESN'T HAPPEN EVERY SPRING —**

- Presenters at numerous sessions of interest to the membership SUBMIT MANUSCRIPTS to *College & University* so that their knowledge and experience and insights may be shared with those registrars and admissions professionals who perforce had to miss certain sessions OR who could not attend at all. Accordingly, *C&U* carries the following advertisement, the first such in our journal’s history.

---

**HELP WANTED**

*College & University*, a juried educational journal, solicits manuscripts from registrars, directors of admissions, and allied professionals. Articles submitted must be of interest to these professions. They may be theoretical, statistical, or practical. They need not be long. Editorial assistance is available. Monetary reward is minimal, but the double satisfaction of becoming a published author AND making a contribution to the field are immeasurable.

Phone the Editor at (414) 288-1901 for additional details.

Or send manuscript to:

Roman S. Gawkoski,  
Editor, *College & University*  
Office of the Registrar  
Marquette University  
P.O. Box 1881  
Milwaukee, WI 53201-1881

---

Roman S. Gawkoski, Editor
Students of Color and Doctoral Programs: Factors Influencing the Application Decision in Higher Education Administration

Michael C. Poock

Graduate education has received a great deal of attention over the last several years, with perhaps the greatest focus being placed upon the institutional choice of prospective students. Unlike the voluminous research on undergraduate college choice during the 1970s and 1980s, the institutional choice of graduate students received most of its attention during the 1990s (e.g., Fox 1992; Kallio 1995; Olson 1992; Webb and Allen 1994; Weiler 1993). Such studies tended to focus on a variety of academic programs at a single institution or a single program at a handful of institutions. Additionally, they tended to examine only master’s students, or master’s and doctoral students with no delineation of the two. Indeed, studies which address the institutional choice of doctoral students are rare (e.g., Talbot, Maier, and Rushlau 1996).

More rare, still, is the examination of the role of ethnicity in the institutional choice of doctoral students. Clearly, ethnicity has received attention in earlier studies on graduate enrollment (Cowell 1985; Nettiles 1990; Olson 1992; Penaloza and Gilly 1991; Fruitt and Isaac 1985). However, the role of ethnicity in the institutional choice of doctoral students is clearly lacking in the literature. The importance of understanding why students of color choose to apply to one institution over another cannot be overstated, as institutions are facing increasing pressure to increase the enrollment of ethnic minority students (Bateman and Hossler 1996).

A cursory review of the data provided by the National Center for Education Statistics (1996) on graduate degrees conferred provides clear evidence of the limited representation of students of color in doctoral programs. Of the doctoral degrees conferred during the 1976-1977 academic year (regardless of discipline), 3.8 percent were earned by African Americans, 1.6 percent by Hispanics, and 2.0 percent by Asian Americans. During the 1984-1985 academic year, the percentage for African Americans dropped to 3.6 percent, but increased for Hispanics and Asian Americans to 2.1 percent and 3.4 percent, respectively. African Americans continued to earn relatively fewer doctorates during the 1993-1994 academic year with just 3.2 percent, while Hispanics remained stagnant at 2.1 percent, and Asian Americans increased to 4.7 percent.

In the academic field of education in general, African Americans fared far better. During the 1993-1994 academic year, African Americans earned 7.5 percent of the doctoral degrees conferred. During this same period, Hispanics earned 2.9 percent of the degrees, and Asian Americans earned just 2.2 percent of the doctorates. However, little is known about specific disciplines within the field of education, and higher education administration is no exception.

The purpose of this study, therefore, is to examine factors influencing the application decision of students of color in doctoral programs of higher education administration. That is, why do students of color apply to specific doctoral programs in higher education administration?

Theoretical Base

For institutions to increase the number of students of color in doctoral programs, it is first necessary to understand why such students choose to apply to specific programs. Although a number of models describe the college choice process at the undergraduate level (for example, Chapman 1981; Litten 1982), this is clearly not the case at the graduate level. As Kallio (1993) points out, “there has not evolved at the graduate level working hypotheses or theories regarding the enrollment decision process.” Therefore, models that delineate the choice process of undergraduates into many discrete stages may not be applicable to graduate enrollment decisions. More applicable would be a model that combines into a few broad stages the common elements from a variety of theories of enrollment choice.

An often utilized model (e.g., Treseder 1995; Waters 1992) developed by Hossler and Gallagher in 1987 meets this need, by identifying three basic phases describing institutional selection by students. The first phase, predisposition, identifies individuals’ background and characteristics that are positively correlated with college attendance. That is, students determine if they want to attend college or pursue other options. The second phase, search, outlines the dynamic process whereby students decide whether or not to attend college, and to which col-

Michael Poock is Assistant Director for Academic and Student Affairs in the Graduate School at the University of North Carolina - Chapel Hill.

Vol. 74, No.3
leges they should apply if they indeed choose to pursue a postsecondary education. It is during this phase that greater interaction between students and institutions of higher education begins to occur. The final phase, choice, is the climax of the college selection process. It is during this phase that institutions increase their communication with students and stress other courtship activities, while students evaluate their options and determine which specific institutions to attend.

The author chose to base this study on Hossler and Gallagher's (1987) model. Specifically, this study addressed the second phase of the selection process: the factors influencing doctoral students' decisions to apply to their current institutions.

The Instrument

Data were collected using The Program Choice Questionnaire. This instrument was developed from relevant literature, refined using focus groups with doctoral students, and pilot tested with numerous graduate students. A five-point Likert scale was used, with the format and design of the questionnaire based on guidelines offered by Fowler (1995) and Schuman and Presser (1996).

Validity of this instrument was established through two means: face validity and content validity. For the former, two researchers with experience in survey research reviewed the instrument. The latter was achieved through a modified focus group format using doctoral students in higher education administration.

Reliability was established by a test-retest, which is advocated for single question measures of factors such as this study presents (de Vaus 1990). A two-week interval was used between administration of the questionnaires to doctoral students, a time period recommended by de Vaus (1990) and Litwin (1995). The result was a correlation coefficient of .9923.

Methodology

Subjects selected for this study were doctoral students who were enrolled in programs of higher education administration for at least one term between fall 1995 and fall 1996, inclusive. Data were collected, as previously noted, using The Program Choice Questionnaire.

Program coordinators at twenty-four randomly selected universities throughout the country were asked to assist in the distribution of The Program Choice Questionnaire. Seven of the program coordinators sent the names and addresses of their doctoral students directly to the researcher. The remaining seventeen program coordinators would not release that information, but did agree to receive questionnaires in unaddressed sealed envelopes and pass them on to their doctoral students through departmental mailboxes, handing them out in class, or attaching mailing labels to the questionnaire envelopes and mailing them directly to the students. Two hundred twenty-five questionnaires were mailed directly to students by the researcher, and 290 were mailed indirectly to students through the program coordinators. Instructions on the instrument asked each recipient to return his/her completed survey directly to the researcher in an accompanying self-addressed stamped envelope.

Of the 515 questionnaires mailed to doctoral students across the United States, 125 were returned from students who were not part of the target population. Excluding these from the study produced a net of 390 possibly valid outstanding questionnaires. Of these, 180 (46 percent of 390) were returned from target population subjects.

Results

Table 1 identifies the rating of the most prominent factors influencing the decision to apply to doctoral programs in higher education administration. Due to space constraints, only factors with means of 3.00 or greater (on a five-point Likert scale) are presented in Table 1.

The factors in Table 1 are fairly self-explanatory. However, there are three themes which warrant attention. First, faculty appear to have a great deal of influence on the decision to apply to a specific program. This includes an overall positive interaction with the faculty—including the admission interview—reputation of the faculty, and the curricular elements over which the faculty have influence (e.g., flexible program requirements, evening classes, diverse course offerings).

Second, people other than faculty also influence the application decision. These include spouse or partner, current students in the program, employers, and alumni. This suggests that a web of influence exists as students make the decision to apply to a program, and that such decisions tend not to occur in a social vacuum.

Third, perceived quality carries a great deal of weight. From the reputation of the program, reputation of the institution, and reputation of the faculty, to the academic accreditations and rigor of the program, respondents appear to be influenced by programs of substance.

The influence of specific factors, however, varies along ethnic lines. That is, while some factors tend to be influential for applicants as a whole, other factors tend to be given more weight when specific racial groups are analyzed. Prior to exploring these results, however, a modification to the data needs to be noted. Data for race were collapsed as to avoid empty cells in the statistical tests. As a result, race was grouped into White, "others," and "Asian."
African American, and non-White/non-African American. The response rate by race follows: White, 139 (77 percent); African American, 22 (12 percent); non-White/non-African American, 19 (11 percent).

As the analysis of variance (ANOVA) results in Table 2 indicate, eighteen factors varied significantly by race. Many of these were highly rated, with means in excess of 3.50. These include:

<table>
<thead>
<tr>
<th>Factors</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible program requirements</td>
<td>4.26</td>
<td>.93</td>
</tr>
<tr>
<td>Availability of evening classes</td>
<td>4.07</td>
<td>1.44</td>
</tr>
<tr>
<td>Ability to continue working in current job</td>
<td>4.06</td>
<td>1.58</td>
</tr>
<tr>
<td>Location (close to home)</td>
<td>4.04</td>
<td>1.39</td>
</tr>
<tr>
<td>Reputation of program</td>
<td>3.96</td>
<td>1.02</td>
</tr>
<tr>
<td>Academic accreditations</td>
<td>3.96</td>
<td>1.13</td>
</tr>
<tr>
<td>Reputation of institution</td>
<td>3.94</td>
<td>.99</td>
</tr>
<tr>
<td>Positive interaction with faculty</td>
<td>3.92</td>
<td>1.31</td>
</tr>
<tr>
<td>Ability to pursue studies as a part-time student</td>
<td>3.86</td>
<td>1.58</td>
</tr>
<tr>
<td>Reputation of faculty</td>
<td>3.85</td>
<td>1.04</td>
</tr>
<tr>
<td>Length of time required to complete program</td>
<td>3.75</td>
<td>1.14</td>
</tr>
<tr>
<td>Input from current professionals or colleagues</td>
<td>3.68</td>
<td>1.25</td>
</tr>
<tr>
<td>Diversity of course offerings</td>
<td>3.68</td>
<td>1.14</td>
</tr>
<tr>
<td>Library facilities and collections</td>
<td>3.55</td>
<td>1.18</td>
</tr>
<tr>
<td>Total credits required</td>
<td>3.51</td>
<td>1.26</td>
</tr>
<tr>
<td>Input from spouse or partner</td>
<td>3.50</td>
<td>1.53</td>
</tr>
<tr>
<td>Geographic region of institution</td>
<td>3.48</td>
<td>1.50</td>
</tr>
<tr>
<td>Rigor of program</td>
<td>3.36</td>
<td>1.14</td>
</tr>
<tr>
<td>Faculty-student ratio</td>
<td>3.34</td>
<td>1.26</td>
</tr>
<tr>
<td>Input from student(s) currently in program</td>
<td>3.28</td>
<td>1.44</td>
</tr>
<tr>
<td>Faculty interview of applicant</td>
<td>3.25</td>
<td>1.43</td>
</tr>
<tr>
<td>Attended as a master’s student</td>
<td>3.25</td>
<td>1.44</td>
</tr>
<tr>
<td>Cost</td>
<td>3.12</td>
<td>1.34</td>
</tr>
<tr>
<td>Access to current technology</td>
<td>3.10</td>
<td>1.33</td>
</tr>
<tr>
<td>Sensitivity to minorities, women, etc.</td>
<td>3.09</td>
<td>1.48</td>
</tr>
<tr>
<td>Input from employer</td>
<td>3.04</td>
<td>1.49</td>
</tr>
<tr>
<td>Flexible entrance requirements</td>
<td>3.01</td>
<td>1.38</td>
</tr>
<tr>
<td>Opportunity for assistantship</td>
<td>3.00</td>
<td>1.68</td>
</tr>
<tr>
<td>Input from alumni</td>
<td>3.00</td>
<td>1.53</td>
</tr>
</tbody>
</table>

1. Academic accreditations (F<sub>2, 170</sub> = 4.56, p < .05)
2. Sensitivity to needs/interests of minorities, women, etc. (F<sub>2, 167</sub> = 10.94, p < .001)
3. Flexible entrance requirements (F<sub>2, 169</sub> = 3.91, p < .05)
4. Rigor of the program (F<sub>2, 166</sub> = 4.80, p < .01)
5. Reputation of the program (F<sub>2, 172</sub> = 4.11, p < .05)
6. Input from alumni (F<sub>2, 158</sub> = 5.60, p < .01)
7. Spouse or partner educational plans (F<sub>2, 88</sub> = 3.93, p < .05).

Newman-Keuls post hoc tests revealed that all significant differences involved students of color rating the factors higher than did White students. Indeed, in no case was there a single significant difference in which White students rated a factor significantly higher than either African American or non-White/non-African American students.

African American students were more heavily influenced than White students with regard to the following highly rated factors: academic accreditations, rigor of the program, and—not surprisingly—sensitivity to the needs/interests of minorities, women, etc. Non-White/non-African American students were more influenced than were White students with regard to flexible entrance requirements, input from alumni, spouse or partner educational plans, and sensitivity to needs/interests of minorities, women, etc.

In contrast to the above, and despite results that were statistically significant,
many other factors were rated only moderately important with means below 3.50. These include, among others, opportunity for internship ($F_{2,170} = 4.41, p < .05$), size of department ($F_{2,107} = 3.07, p < .05$), and affordability of off-campus housing ($F_{2,99} = 4.06, p < .05$). All these moderately important factors also involved students of color rating the factors as more important than did White students: a location that is far from home, availability of graduate student services, opportunity for internship, size of department, and availability of child care. Non-White/non-African American students rated affordability of off-campus housing, friends attending the institution, and friends living in the area significantly higher than did White students.

**Discussion**

From an aggregate perspective, doctoral students in higher education administration were influenced by a number of factors in their decision to apply to their current programs. Overall, the factors receiving the greatest weight were those related to program quality, faculty, and input from spouses/partners, alumni, and students. Yet, such an aggregate view may blur the true picture of the impact of various factors on the application decision of students who are ethnic minorities.

As noted at the outset, appropriate racial representation within the student body is often a concern of program coordinators and faculty in academic programs. Thus, knowing what positively influences students of color can only help program coordinators and faculty in their recruitment. While it is inaccurate to suggest or imply that members of any racial group have inherent similarities, the data do suggest that on average some factors are considered more important for some students of color when compared to White students. A theme apparent from these findings is that people and the acad-

<table>
<thead>
<tr>
<th>Factors</th>
<th>White Mean</th>
<th>White SD</th>
<th>African American Mean</th>
<th>African American SD</th>
<th>Other Mean</th>
<th>Other SD</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic accreditations</td>
<td>3.83</td>
<td>1.14</td>
<td>4.60</td>
<td>1.67</td>
<td>4.06</td>
<td>1.23</td>
<td>4.56*</td>
</tr>
<tr>
<td>Location (far from home)</td>
<td>1.93</td>
<td>1.39</td>
<td>3.27</td>
<td>1.39</td>
<td>2.10</td>
<td>1.10</td>
<td>5.98**</td>
</tr>
<tr>
<td>Graduate student services</td>
<td>2.41</td>
<td>1.32</td>
<td>3.23</td>
<td>1.45</td>
<td>2.61</td>
<td>1.29</td>
<td>3.61*</td>
</tr>
<tr>
<td>Sensitivity to needs/interests of minorities, women, etc.</td>
<td>2.82</td>
<td>1.45</td>
<td>4.18</td>
<td>1.10</td>
<td>3.72</td>
<td>1.32</td>
<td>10.94***</td>
</tr>
<tr>
<td>Flexible entrance requirements</td>
<td>2.89</td>
<td>1.34</td>
<td>3.05</td>
<td>1.47</td>
<td>3.83</td>
<td>1.30</td>
<td>3.91*</td>
</tr>
<tr>
<td>Rigor of the program</td>
<td>3.21</td>
<td>1.13</td>
<td>3.95</td>
<td>1.92</td>
<td>3.67</td>
<td>1.28</td>
<td>4.80**</td>
</tr>
<tr>
<td>Reputation of the program</td>
<td>3.84</td>
<td>1.07</td>
<td>4.41</td>
<td>1.73</td>
<td>4.28</td>
<td>0.75</td>
<td>4.11*</td>
</tr>
<tr>
<td>Opportunity for internship</td>
<td>2.51</td>
<td>1.45</td>
<td>3.33</td>
<td>1.46</td>
<td>3.26</td>
<td>1.66</td>
<td>4.41*</td>
</tr>
<tr>
<td>Size of department</td>
<td>2.78</td>
<td>1.09</td>
<td>3.43</td>
<td>1.21</td>
<td>2.94</td>
<td>1.26</td>
<td>3.07*</td>
</tr>
<tr>
<td>Post-graduate job placement</td>
<td>2.62</td>
<td>1.38</td>
<td>3.60</td>
<td>1.27</td>
<td>2.89</td>
<td>1.53</td>
<td>4.40*</td>
</tr>
<tr>
<td>Campus visit</td>
<td>2.48</td>
<td>1.45</td>
<td>3.05</td>
<td>1.40</td>
<td>3.64</td>
<td>0.84</td>
<td>5.18**</td>
</tr>
<tr>
<td>Input from alumni</td>
<td>2.79</td>
<td>1.53</td>
<td>3.13</td>
<td>1.36</td>
<td>4.23</td>
<td>1.09</td>
<td>5.60**</td>
</tr>
<tr>
<td>Spouse/partner educational plans</td>
<td>2.34</td>
<td>1.65</td>
<td>3.17</td>
<td>1.75</td>
<td>3.78</td>
<td>1.30</td>
<td>3.93*</td>
</tr>
<tr>
<td>Availability of child care</td>
<td>1.64</td>
<td>1.33</td>
<td>3.22</td>
<td>1.92</td>
<td>2.11</td>
<td>1.69</td>
<td>4.63*</td>
</tr>
<tr>
<td>Affordability of off-campus housing</td>
<td>2.15</td>
<td>1.42</td>
<td>2.00</td>
<td>1.70</td>
<td>3.42</td>
<td>1.62</td>
<td>4.06*</td>
</tr>
<tr>
<td>Cost of living in area</td>
<td>2.33</td>
<td>1.45</td>
<td>2.27</td>
<td>1.79</td>
<td>3.57</td>
<td>1.56</td>
<td>4.25*</td>
</tr>
<tr>
<td>Friends attend institution</td>
<td>2.63</td>
<td>1.28</td>
<td>1.75</td>
<td>1.23</td>
<td>2.85</td>
<td>1.21</td>
<td>3.07*</td>
</tr>
<tr>
<td>Friends live in the area</td>
<td>1.97</td>
<td>1.33</td>
<td>1.60</td>
<td>0.91</td>
<td>3.00</td>
<td>1.05</td>
<td>4.53*</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.
emic environment tend to be more important for students of color. Other factors which were rated high, with respective means in excess of 3.50, include input from alumni, spouse or partner educational plans, a campus visit, and an institutional sensitivity to the needs/interests of minorities, women, etc.

This suggests that students of color do not simply evaluate a program based on the academic reputation or friendliness of the faculty alone. Rather, it appears that these students also choose to investigate the environment. This appears to be done via talking to other students who have been through the program, visiting the campus firsthand, and ensuring that the institution is aware of—and sensitive to—their needs.

Armed with this information, program coordinators and faculty in higher education administration programs can take specific steps to assist students of color in their efforts to investigate the academic environment. In addition to incorporating relevant factors rated high by the respondents as a whole (such as positive interaction with faculty), program coordinators and faculty should take the initiative to develop an alumni network and an organized campus visitation schedule so the burden for organizing these does not fall to the prospective students. Additionally, program coordinators and faculty can include the spouses or partners in these activities, and assist them in obtaining information relevant to their own educational plans. Since students of color tend to examine the academic environment through a web of social influence, any effort on the part of the program coordinator and faculty in a higher education administration program to streamline these efforts—and ensure they are available—would likely increase the probability that prospective students of color apply to their doctoral program.

**Recommendations for Future Research**

Understanding why students apply to specific doctoral programs is of increasing importance to departments of higher education administration specifically, and to universities as a whole. This point was noted by the former Dean of Graduate Studies at the University of Michigan who argued that to effectively plan for the future needs of academic programs and graduate students, "We need to know more about students’ reasons for applying to the University of Michigan and for choosing to enroll or not once admitted" (D’Arms cited in Kallio 1993). Given the underrepresentation of students of color in doctoral programs, and the related pressures to increase their enrollment, understanding the factors influencing the application decisions of students of color becomes ever more paramount. While it is hoped that this study took a fundamental step toward that understanding, there is still much more that can be done. As a result, the following are recommendations for future research.

First, replicate this study while increasing the number of students from various racial groups. Since the data related to race were collapsed into three categories because of the limited number of students of color in this study, few conclusions could be drawn concerning students who were neither White nor African American. Greater representation by students of all races would help paint a clearer picture of the institutional choice of students of color.

Second, conduct a longitudinal study with a focus on retention. It would be extremely helpful to learn if—and to what degree—factors related to the decision to apply to an institution have an influence on retention rates of students of color. Indeed, the best recruitment efforts amount to naught if the students do not remain with the program.

Finally, examine the recruitment practices of program coordinators and faculty in doctoral programs in higher education administration. While studies of doctoral program recruitment efforts exist, there is a dearth of information regarding higher education administration programs in general, and recruitment practices towards students of color in particular.

**Conclusion**

Understanding why students of color apply to specific doctoral programs in higher education administration is of increasing importance to both academic departments and universities as a whole. The impetus for this is easily seen by the underrepresentation of students of color combined with the increased institutional pressure to increase enrollment of ethnic minorities. While this study examined the identification of factors that influence students of color to apply to specific doctoral programs in higher education administration, this area is in need of further exploration.

**Limitations**

A clear limitation to this study is the sample size of the ethnic subpopulations. While the proportion of African American and non-White/non-African American respondents exceeds that found in the general population, the limited number of such doctoral students makes obtaining a larger sample size problematic. Statistically, a smaller sample increases sampling error, yet aggregating respondents on the basis of race may blur any distinction among racial groups. As a result, any unique differences among Native Americans or other non-White/non-African American racial groups cannot be further identified in this study.
References


Employing Quantitative Models of a Qualitative Admissions Process: Uncovering Hidden Rules, Saving Time, and Reducing Bias

Philip M. Sadler and James K. Hammerman

Abstract

Admission into highly selective institutions of higher education must often rely on the judgments of panelists who first independently assign applicants to ordinal categorical or numerical scales. Such initial decisions are then used in a multi-step winnowing process of discussion and group decision. Our study quantitatively models this inherently subjective process using five years of graduate admissions data encompassing 592 candidates and 72 raters. Logistic regression models are found to be well-fitting and parsimonious. They allow for an analysis of the contribution of each stage of the process of which the initial, blind rating of candidates accounts for roughly 40 percent of the variance in admissions. Our models also reveal the tacit rules governing admissions decisions, including the relative power of each classification category and implicit thresholds. We find that extended discussion and deliberation phases requiring the entire committee can be of limited productivity when inter-rater agreement is high. Both rater bias and the use of linear scaling are explored as threats to fairness and remedies are examined to make admissions decisions more equitable and efficient.

Introduction

Since the first statutes governing admissions were drawn up by Harvard College, admission into American higher education has only become more complicated. In 1642, knowledge of Latin and Greek insured entry into this struggling college. By 1734, examinations were required. In 1807, mathematics, geography, and proof of "good moral character" were added. By 1875, knowledge of a foreign language, English literature, physical science, geometry, and algebra boosted the requirements further (Broome 1903). Conditions now are certainly quite different from 350 years ago when Harvard was constantly under the threat of financial collapse. Competition has grown to the extent that at many highly selective schools, entrance is only offered to 10 percent of applicants. A similar situation is mirrored by many graduate programs.

While many colleges and universities use a formulaic approach in combining standardized test scores and academic achievement to select an entering cohort, others use a mix of measures, often relying on evidence not prone to exact quantitative encoding. Ivy League and other highly competitive schools generally employ a multi-stage process for undergraduate admissions in which applications are individually rated by admissions officers and the pool winnowed by committee (Hernández 1997).

Graduate admissions can follow a very different process than that for undergraduates. Undergraduate admission most commonly is carried out by admissions counselors with little involvement by faculty. When an interview is part of the admissions process, alumni often conduct it far from campus. By contrast, graduate admissions committees are often made up of junior and senior faculty members, an admissions office representative, and, in some schools, graduate students (Sacks 1978) because selecting doctoral students is a long term investment, often viewed as an opportunity to match the resources and

Philip M. Sadler is Assistant Professor of Education at the Harvard University Graduate School of Education. For seven years he has served on the Doctoral Admissions Committee for Learning and Teaching, for the last two as its chairman. Dr. Sadler also heads the Science Education Department at the Harvard-Smithsonian Center for Astrophysics (CfA). He delights in using quantitative tools to understand qualitative issues, such as admissions decisionmaking, assessing students' conceptual understanding in science, and the engagement of students in research activities. He has developed curricula and software for teaching science in conjunction with leading professional development experiences for precollege teachers.

James K. Hammerman is a teacher and mathematics teacher educator and is currently a doctoral candidate in Learning and Teaching at Harvard's Graduate School of Education. He has written a variety of mathematics, geography, and professional development curricula, and has taught and consulted with teachers in several grant-funded programs at Education Development Center (EDC) and SummerMath for Teachers. His research interests focus on adult developmental issues associated with transformational teacher professional development programs.

Vol. 74, No. 3
needs of a school with an applicant's interests and talents, impressions and intuitions must substitute for the comfort of numerical scores. We have studied the Harvard University Graduate School of Education's Doctoral Program in Learning and Teaching over a five year period to characterize a three-stage admissions process that relies heavily on judgments of quality based in complex data.

Prior work using a researcher's systematic tools to examine admissions has taken the form of examining the effective use of quantitative selection criteria in predicting later performance in graduate school. Academic success in graduate programs has been examined utilizing measures of undergraduate college caliber, college transcripts, and standardized test scores (MAT, GRE, MCAT, GMAT). Medical (Hall 1992), nursing (Rhodes 1994), business (Fisher 1990; Graham 1991; Zwick 1993), physical therapy, (Seymour and Gramet 1995) and veterinary schools (Stuck 1990) have used such data to predict graduate school grades and, in addition, scores on post-graduate professional examinations (Fletcher 1989; Mitchell 1990). Studies in education have been limited to undergraduate admissions programs (Fletcher 1989; Smith and Pratt 1996).

Other information embedded in students' application portfolios has been examined as predictive of later performance. Parental socio-economic status, work experience, applicant reports of gender and ethnic status, and expectations of the faculty were used to predict subsequent academic self-confidence and self-sufficiency in engineering and physical science programs (Einarson 1996). The Dartmouth Medical School found that the quality of admission interviews correlated positively with dean's letter ratings four years later. Interview scores outperformed MCAT and undergraduate GPA in predicting scores on professional medical exams (Hall and Bailey 1992). Innovative admission changes including portfolio assessment, two-tier admissions, and group interviews have also been found to be more effective models for predicting the success of women, minorities, and older students (Hagedorn 1996).

Processes that rely more heavily on such embedded information are much more labor intensive than those that make decisions based only upon quantitative measures. Finding ways to save time while using more holistic methods, without compromising the quality of the process, is a reasonable goal (King, Bruce, and Gilligan 1993; Nagy, Cotter, et al. 1993). Admissions is just one example of a high-stakes assessment prone to subjectivity (e.g., portfolios, essays, performance assessments) for which the issues of decision-making and agreement between judges has become increasingly important (Burry-Stock, Shaw, et al. 1996).

Few have studied the admissions process itself. Admissions meetings are typically shrouded in secrecy. Applicants have little idea of the path their file follows. The origins of admissions procedures are lost to the past; in the present we simply do what was done previously. The Educational Administration Leadership Program at The University of Texas at Austin is one program that has critically examined its methods. Interviews, oral presentations, and writing analyses were compared with more quantitative measures as predictors for admission. Use of these more subjective measures has increased minority matriculation and resulted in higher student placement rates at the program's end (Veir 1990).

Others have studied the utility of multiple regression models for veterinary school admission. One study found that special attention must be paid to validating models developed by analyzing prior cohorts when applied to current applicants (Stuck 1990). Another study reported an attempt to systematize the selection of candidates by identifying those criteria that appear to best predict which candidates are actually selected by admissions committees (King, et al. 1993). In future years, using such criteria as a filter, a more streamlined process can be developed that may be useful as a step to narrow to a pool of the most highly qualified candidates.

We were also interested in increasing admissions effectiveness while guarding against employing overly deterministic methods. Graduate admissions should use tools that increase the reliability and efficiency of decisions, while not compromising the complex human aspects of selecting students from a large group of applicants. Another way to think of this is not to limit the criteria by which we judge applicants to those criteria more susceptible to quantification (e.g., test scores, GPA, years of experience), but to systematize and reduce error in the measurement of more subjective traits which are deemed critical (e.g., potential for leadership, fit with our offerings). Quantitative tools can also help to monitor and assess biases.

In this five-year study, the admissions process ran relatively unchanged for the first four years. For its final year (1997), several modifications were made in the hope of maintaining or improving the quality of decisions, while reducing the workload of committee members. Among these is an attempt to recognize and use historical patterns. It is important to realize that the time required to select a cohort is affected by the degree to which fellow committee members' independent decisions are viewed as reliable proxies for one's own reading and discussion. This analysis helps in understanding whether or not such changes have merit.
Harvard’s Admissions Process

Harvard’s doctoral program in Learning and Teaching typically admits 17 candidates out of a varying field of 100-150 applicants using a committee of 12 to 18 members. Membership on this committee requires a sizable time commitment to read files and attend meetings. Members report that they can read files at a rate of two to four per hour. The committee meets as a group for approximately twenty-four hours, although in the last year this time has been reduced by half. The admissions process is organized into three discrete stages. First, individual members independently rate candidates after reading their admissions folders. Second, each candidate is presented to and discussed by the committee and rated again through group consensus. Third, candidates are compared and the final disposition of each is decided.

Criteria for rating applicants are discussed at the first meeting of the admissions committee each year. Conventional measures of prior success are submitted along with each application. These include:

1. A personal statement tracing educational and professional development and future goals.
2. Three recommendations from colleagues, supervisors, or professors.
3. Undergraduate and graduate transcripts, and test scores.¹

These sources are read for clues to a candidate’s potential for educational leadership, depth of educational ideas, match with the program’s strengths and resources, and motivation for embarking on a doctoral program. In return for their submission, candidates expect fair treatment including a full review of their application.

Committee Formation

Within this graduate school, it has long been the tradition that admissions committees are composed of both faculty and students. Faculty volunteers. Students apply for a slot, are interviewed, and receive an explanation of the procedures and time commitment. If there are more students than slots available, the chairperson selects who will be a member with an eye to increasing the ethnic and gender diversity of the entire committee.

An initial meeting orients new committee members to procedures and responsibilities. The criteria that past members have used for evaluation are discussed. Timelines and meeting dates for the future are scheduled. Committee members are asked to read between twenty-five and forty-five folders, with two faculty and two student readers per candidate.

A Three Stage Admissions Process

In the first stage of the process, four committee members read a candidate’s folder and rate it as Admit (A), Tend-to-Admit (TA), Tend-to-Deny (TD), or Deny (D). Raters are blind to others’ ratings.² Candidates are then added to a docket for discussion by the full committee. Through 1996, at each admissions meeting, every candidate was presented (in alphabetical order) by the member who rated them highest, discussed, and then placed into one of four intermediate groupings. In order, the group decisions are Group Admit (GA), Group High Pool (GH), Group Pool (GP), or Group Deny (GD). At a final meeting a cohort is chosen considering group ratings, initial ratings, and any other pertinent information brought forward by the readers.

With multiple readers for each candidate, the admissions process is assumed to be fair since any systematic bias in one reader’s rating will tend to average out over all readers. The statistical bias of readers has not previously been raised as a threat to fairness of this process. However, rater stringency has been an issue in other studies of the quantification of qualitative assessments. Post-hoc statistical fixes have been suggested to remedy systematic biases (Cason and Cason 1989). Yet, most of the literature on inter-rater reliability does not apply to admissions, since there is no “true score” with which rater estimates may be compared.³

Changes Over Time

The process of selecting an entering cohort in our doctoral program remained consistent during the period from 1992-1996, although several faculty members felt that too much time was spent in meetings (approximately eight two-hour meetings and an eight-hour final meeting) for the amount of time spent reading folders (ten to fifteen hours). As a result, several changes were implemented for the 1997 selection procedure.

• All folders were read prior to any group decisions.⁴

¹Either the Graduate Record Exam (GRE) or the Miller Analogy Test (MAT). Foreign students must take the Test Of English as a Foreign Language (TOEFL).
²The only exception to the “rule of four readers” is that if candidates receive three low ratings (TD or D), a fourth reading will not be made (instituted in 1996).
³In a “guess your weight” carnival sideshow, ratings can be compared to a true measure. In admission, there is no true measure of candidate quality, so we must explore inter-rater agreement.
⁴Except for a few in an initial “calibration” meeting, so that students and new faculty could become familiar with the rating process.
• A single meeting was held to make group and final decisions.

• Prior to discussion, all candidates were listed in order of initial ratings.

• The applicants were discussed starting from the highest rated.

• Discussion of any applicant below two Tend-to-Admits was delayed until after all higher rated candidates were discussed.5

These decisions made possible a general ranking and the establishment of a cut-off under which candidates would not be discussed. Starting with a discussion of the highest rated candidates, members had a standard with which to compare. With only 18 openings for a list of 140 applicants, it quickly became evident that to accept any candidate lower than the eighteenth position, someone above must be argued against. These new guidelines reduced the meeting time by half: a one-hour orientation, the two-hour “calibration” discussion, and a ten-hour final meeting.

Methods

Data Collection

Data were collected for five years, recoding it so that the identity of both applicant and rater remain anonymous. The record identified initial ratings, pool rating, and final acceptance or denial for each candidate. In total, 592 candidates vied for 99 slots. Committee members performed 2,364 readings of folders.

Analyses

The goal of our analyses has been to answer these questions:

• Can the admissions process be accurately modeled using past data?

• What is the contribution of each stage of admission to the final decision?

• To what degree do committee members’ ratings agree?

• Do raters exhibit any systematic bias that compromises fairness?

• What are the implications of our findings for other admissions programs?

To answer these questions, we begin by calculating descriptive statistics for each of five years. These include summaries of ratings by individuals and by raters as a whole. Each stage of the admissions process has been examined using the final admission decisions as a guide. The admissions process is modeled by regression equations, of which the logistic form is most applicable. To ascertain bias, committee members’ ratings are compared. Several measures of inter-rater bias are used to calculate systematic deviations of raters’ scores from that of the group. The magnitude of agreement between raters is estimated using the conventional Spearman-Brown coefficient (S-B) and by the Rater-Agreement-Index (RAI).

Regression Models to Explain Variance

In this study, regression models are used to study the information flow within the admissions process itself. Logistic regression models are constructed to examine the variance explained by the different stages of the process. The more variance explained, the better the model is in predicting who will ultimately be admitted or denied. Logistic regression makes far more sense for modeling admissions than does linear regression, since admission can be thought of as a dichotomous outcome which has a value of either 0 or 1. Logistic models are used to calculate probabilities (also ranging from 0 to 1) of admission outcomes using variables such as initial ratings.6

Models are built first from the initial ratings (A, TA, TD, D) and then from the group ratings (G-A, G-H, G-P, G-D) to predict the final decisions of the committee (see Equation 1).7 For initial ratings, variables can have a value of 0, 1, 2, 3, or 4 representing the number of votes in each category among the four raters. For group decisions, values are 1 in the rated category or 0 otherwise, because candidates only get one group rating. Ratings are normalized for candidates who received fewer or more than four readings.8 We seek to build a model with the fewest terms yet still explains the most variance in the final admissions results.

These logistic models are used to estimate the fraction of the decision-making

---

5Candidates below two tend-to-admits are not earmarked for discussion unless one reader feels especially strongly about their candidacy.

6Linear regression can produce nonsensical results with predicted probabilities of greater than 1 or less than 0.

7Only three variables are needed to define the model.

8When a particular faculty member’s interests are presented in the candidate’s personal statement, occasionally an additional reading is made.
process represented by each stage in the committee's work. The pseudo-R^2 statistic (referred to as R^2 in this study) calculates the variance explained by the logistic model (Hosmer and Lemeshow 1989; Knoké and Bohrnstedt 1982). Initial ratings are a first step in winnowing the applicant pool. Group decisions are the second. The final meeting where the cohort is chosen can be thought of as explaining all of the remaining variance. Each step should contribute additional explained variance and the inclusion of the results from each step should produce a model which more accurately predicts the ultimate outcome.

Regression models are also fit to determine the degree of improvement in proportion to explained variance by accounting for rater bias. Accounting for any systematic rater error should reduce the unexplained error and increase the proportion of variance explained by the model. If no such increase results, accounting for such bias is unwarranted.

A logistic regression model is also developed by pooling the first four sets of data to test the fifth year of data. In this way, a model can be validated; the fitted model is used to predict the outcome for future subjects. The probability estimated by models can then be fitted to observed values in any number of unique partitions of data using Tsiatis' variation of the Hosmer-Lemeshow goodness-of-fit statistic, \( \hat{C} \) (Hosmer and Lemeshow 1989). \( \hat{C} \) is evaluated using a Pearson Chi-square statistic with g - 1 degrees of freedom (Equation 2).

### Inter-rater Bias

Committee members rate applicants differently. Some are severe; others are more lenient. The random assignment of applicants to raters does not eliminate the possibility that a candidate can draw a set of reviewers that, taken together, may be overly harsh or easy. If each committee member rates every folder, systematic bias in raters would both be simple to establish and cancel out. With just a subset reading each folder, inter-rater bias can only be estimated, since only four raters look at each folder out of a larger committee. Bias in ratings is explored by looking at raters' decisions in two ways.

For the purpose of assessing rater bias, the four rating categories are assigned numeric values from 1 to 4 (A=4, TA=3, TD=2, D=1), so that the calculated bias can be compared in units of difference between adjacent categories. Using this method, average ratings are calculated for each rater (or judge) over all the candidates whom they rate.

Subtracting the average of all judges' mean ratings from an individual rater's mean assigns a handicap to each rater (Equation 3). These handicaps can be compared to each other and to the overall average rating. When a handicap is negative, it means a rater tends to be more stringent than others. When it is positive, the rater is more lenient. The handicap can then be subtracted from the rater's rating of a candidate to attempt to equalize bias.

Secondly, even though folders are assigned pretty much at random, it is possible for raters to draw a subset of candidates that, as a group, are stronger or weaker than average.\(^9\) The calculation of a handicap in the fashion described in Equation 3 does not account for dissimilar pools. Since there are four ratings for each committee, bias in raters would both be simple to establish and cancel out. With just a subset reading each folder, inter-rater bias can only be estimated, since only four raters look at each folder out of a larger committee. Bias in ratings is explored by looking at raters' decisions in two ways. Subtracting the average of all judges' mean ratings from an individual rater's mean assigns a handicap to each rater (Equation 3). These handicaps can be compared to each other and to the overall average rating. When a handicap is negative, it means a rater tends to be more stringent than others. When it is positive, the rater is more lenient. The handicap can then be subtracted from the rater's rating of a candidate to attempt to equalize bias.

\(^{10}\)This problem was noted by our colleague, Donald Oliver, who suggested that a more sophisticated form of analysis could control this effort.
each candidate, comparisons between an individual rater and the three other colleagues can be made and averaged over all candidates rated by that rater. Using this Handicap-by-case, each rater is compared on a case-by-case basis to only the other three readers for each candidate (Equation 4).

These handicaps can be studied by examining their distribution about zero. Either of these two handicaps can be subtracted from the ratings of applicants in an attempt to "correct" raters' systematic bias.

Inter-rater Agreement

Although we have presented two measures of inter-rater bias, it remains to establish the degree to which our raters agree overall. Inter-rater agreement can be calculated in several ways. The Kappa Statistic is used to assess agreement between judges for nominal data (Fleiss and Cohen 1973). Since our ratings are ordinal, a ratio of variances by case and by judge can be calculated using the Spearman-Brown reliability formula (Winter 1962).

Some researchers argue that reliability is a quite different measure than rater agreement and that inter-rater agreement must be calculated on a case-by-case and judge-by-judge basis. The Rater-Agreement-Index (RAI) is calculated by comparing and summing the difference between individual judge's ratings and the average rating of others who rated the same candidate, similar to the handicap-by-case (Burry-Stock, et al. 1996). When divided by the appropriate degrees of freedom (taking into account the number of judges, candidates, and rating categories), an index with a value from 0.00 to 1.00 characterizes the agreement between raters. This is similar to the calculation of the handicap-by-case calculated above but takes into account the number of degrees of freedom (Equation 5).

Results

Descriptive Statistics

This graduate program is highly selective, with a yearly acceptance rate from 13 percent to 21 percent. Although committee members read a large number of candidates and admission committee.

Table 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Admit</td>
<td>0.15</td>
<td>0.18</td>
<td>0.14</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>Tend-to-Admit</td>
<td>0.29</td>
<td>0.33</td>
<td>0.33</td>
<td>0.26</td>
<td>0.27</td>
</tr>
<tr>
<td>Tend-to-Deny</td>
<td>0.26</td>
<td>0.28</td>
<td>0.32</td>
<td>0.39</td>
<td>0.33</td>
</tr>
<tr>
<td>Deny</td>
<td>0.29</td>
<td>0.21</td>
<td>0.22</td>
<td>0.20</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 1. Descriptive Statistics for 5 Years of Admissions. Fluctuations in the number of candidates and committee members has had little impact on the initial ratings of candidates.
of applicant folders, thirty-five in an average year, few that they read will actually be admitted (in the average year, only six). Appropriately, members do appear to reserve the Admit rating for a few candidates (see Table 1). Pool ratings are even more selective with an average of only 10 percent receiving a Group-Admit rating. The number of candidates has fluctuated over the period of the study, while the number of openings has remained stable. The number of reviewers has changed as well. The distribution of initial ratings (shown in Figure 1) is mono-modal with the Tend-to-Deny category being the most common, not the Deny category. This is somewhat surprising in that the admissions process ultimately selects only one out of seven candidates from the pool; one would expect the Deny category to be the largest. Few applicants initially rated in either of these two lowest categories by any raters are actually accepted into the program. The Deny or Tend-to-Deny ratings appear not to distinguish in any substantive fashion between candidates. In practice, raters appear to assign to the Deny category applicants who are unsuited to graduate work, while Tend-to-Deny is assigned to those who have some major weakness. The distinction is essentially meaningless since it appears that only students with strong credentials are ever accepted.

Patterns are uncovered in the eventual admission outcome by grouping the two highest initial ratings of Admit and Tend-to-Admit into fifteen discrete combinations for the five-year study.\(^\text{11}\) Applicants with four Admit ratings have always gained admission to the program, while those with fewer than two Admits or Tend-to-Admits have never gained entry (shown in Figure 2). Without at least one strong advocate on the committee, the case for admission appears never to be endorsed forcefully enough to gain admission.

---

\(^{11}\) The combined categories of A and TA are: 4A, 3A 1TA, 3A 0TA, 2A 2TA, 2A 1TA, 2A 0TA, 1A 3TA, 1A 2TA, 1A 1TA, 1A 0TA, 0A 4TA, 0A 3TA, 0A 2TA, 0A 1TA, and 0A 0TA.
Group ratings always result from the deliberation of the full committee. These four categories are used more selectively than initial ratings, with more than two-thirds of applicants being placed in the Group Deny category (see Figure 3). The Group-Pool, Group-High-Pool, and Group-Admit are roughly the same size, averaging 11 percent of applicants in each. Ninety-seven percent of candidates with a certain rating to be admitted when presented as odds-ratios, Group-High-Pool and Group-Deny category (see Figure 3). The Tend-to-Admit categories.

Table 3 presents the log-likelihood coefficients converted to odds-ratios for easier interpretation. One can see that in each year, an initial rating of Admit leads to a higher probability of admission than does a Tend-to-Admit rating. The fraction of total variance accounted for by initial ratings alone is large, ranging from 0.334 to 0.425. This first stage in the admissions process makes great headway toward discriminating between those who are ultimately admitted and those who are not.

We also constructed logistic models to explain the probability of admission based upon the second stage of the admissions process, the group-rating. For the group-rating model, only G-A and G-HP categories are significant and included in the final model for each year. Combining data from the first four years of the study allows the additional parameter, Group-Pool, to be added to the 2-parameter (Group-Admit and Group-High-Pool) regression equation at the p ≤ 0.01 level, although this parameter is not significant for any individual year alone.

The proportion of variance accounted for by each stage in the admissions process is presented in Figure 4. Somewhat over a third (0.33 to 0.42) of the differentiation needed in choosing a final cohort is contained in the initial stage. In each year, the second admissions stage, group-rating, accounts for more variance than the initial ratings. The average increase in the explained variance is 0.171. It is reasonable to assume that these increases are indicative that the group-rating process adds incrementally less measurable information to the admissions process than does the first stage, although it may also serve other purposes within the committee—group building, further calibration, or working out the overall character of the cohort. One would hope that the group discussion and assignment of candidates to new categories would account for much more variance in outcome. Each year the largest fraction of variance is accounted for in the third stage of deliberation, the final decision.

Measures of Goodness of Fit

Just how well does a multiple logistic model built from only two variables fit the data? Very well, indeed. The conditional probability of admission predicted by the model has been compared to the actual admission results for the five years of the study. Partitioning the data into fifteen discrete categories represents all the combinations of Admit and Tend-to-Admit categories. These data have been plotted (see Figure 5) comparing the fraction accepted in that category with the fraction of all applicants receiving higher ratings.

The logistic model fits the data well. Bars of ±1 standard error mark the expected range in conditional probability, based on the number of applicants in each category. All the data are seen to be within one standard error of the model probability, save 1A 1TA (fifty-three cases) and 3A 0TA (three cases). The goodness-of-fit of the model to this graph of conditional probabilities, as calculated by a least-squares standardized proportional reduction in error, is quite high, $R^2 = 0.938$. A large fraction of the variation in probability by category is captured by the logistic regression model. The Hosmer-Lemeshow goodness-of-fit statistic computed from the actual and predicted frequencies for these four years of data show.

---

12So, for example, a candidate in 1993 with one Admit rating was 5.66 times as likely to be admitted than one with a TD or D rating. A candidate in 1993 with an Admit rating was 5.66/3.16 = 1.79 times as likely to be admitted as one with a Tend-to-Admit rating.
Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-4.703***</td>
<td>-5.476***</td>
<td>-5.283***</td>
<td>-9.023***</td>
<td>-5.120***</td>
<td>-6.061***</td>
</tr>
<tr>
<td>Admit</td>
<td>1.965***</td>
<td>1.848***</td>
<td>1.663***</td>
<td>2.839***</td>
<td>1.809***</td>
<td>2.150***</td>
</tr>
<tr>
<td>Tend-to-Admit</td>
<td>0.760*</td>
<td>1.259*</td>
<td>1.297*</td>
<td>2.743***</td>
<td>1.194***</td>
<td>1.496*</td>
</tr>
<tr>
<td>Tend-to-Deny</td>
<td>-0.287</td>
<td>0.173</td>
<td>0.003</td>
<td>0.499</td>
<td>0.066</td>
<td>-0.519</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.396</td>
<td>0.422</td>
<td>0.420</td>
<td>0.317</td>
<td>0.426</td>
<td>0.408</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-5.277***</td>
<td>-5.006***</td>
<td>-5.275***</td>
<td>-7.285***</td>
<td>-5.290***</td>
<td>-7.346***</td>
</tr>
<tr>
<td>Admit</td>
<td>2.123***</td>
<td>1.733***</td>
<td>1.661***</td>
<td>2.421***</td>
<td>1.851***</td>
<td>2.457***</td>
</tr>
<tr>
<td>Tend-to-Admit</td>
<td>0.884**</td>
<td>1.149***</td>
<td>1.295***</td>
<td>2.278***</td>
<td>1.234***</td>
<td>1.798***</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.395</td>
<td>0.423</td>
<td>0.420</td>
<td>0.334</td>
<td>0.425</td>
<td>0.410</td>
</tr>
</tbody>
</table>

Table 2. Logistic Models from Initial Ratings. Beta coefficients are for logistic models predicting admission from combined initial rating categories. The top table is a 3-parameter model incorporating the number of Admit, Tend-to-Admit and Tend-to-Deny ratings. The bottom table’s model uses only Admit and Tend-to-Admit ratings. $R^2$ is an $R^2$ type statistic modified for logistic regression. Asterisks represent the statistical significance of the term *** $p \leq .001$, ** $p \leq .01$, and * $p \leq .05$.

Table 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds-Ratio of A:D or TD</td>
<td>8.98</td>
<td>5.66</td>
<td>5.03</td>
<td>8.65</td>
<td>6.37</td>
<td>11.50</td>
</tr>
<tr>
<td>Odds-Ratio of TA:D or TD</td>
<td>2.84</td>
<td>3.16</td>
<td>3.60</td>
<td>7.80</td>
<td>3.43</td>
<td>6.20</td>
</tr>
<tr>
<td>Odds-Ratio of A:TA</td>
<td>3.16</td>
<td>1.79</td>
<td>1.40</td>
<td>1.11</td>
<td>1.85</td>
<td>1.85</td>
</tr>
<tr>
<td>variance accounted for by initial rating</td>
<td>0.40</td>
<td>0.42</td>
<td>0.42</td>
<td>0.33</td>
<td>0.42</td>
<td>0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds-Ratio of GA:GD or GP</td>
<td>3E+38</td>
<td>145</td>
<td>213</td>
<td>1E+15</td>
<td>399</td>
<td>2E+6</td>
</tr>
<tr>
<td>Odds-Ratio of GH:GD or GP</td>
<td>26</td>
<td>16</td>
<td>24</td>
<td>24</td>
<td>15</td>
<td>773</td>
</tr>
<tr>
<td>Odds-Ratio of GP:GD</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>3</td>
<td>---</td>
</tr>
<tr>
<td>variance accounted for by group rating</td>
<td>0.57</td>
<td>0.54</td>
<td>0.48</td>
<td>0.56</td>
<td>0.67</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 3. Odds-ratios and Variance explained by Logistic Regression Models. Regression coefficients are presented as odds-ratios for easier interpretation. Stage 1 accounts for about forty percent of the variability in outcome. Stage 2 explains somewhat more for each individual year except for 1997. The first four years of Stage 1 data expressed as the 15 categorical combinations of Admit and Tend-to-Admit are well fitted by a two-parameter regression equation, where $R^2 = 0.947$. The first four years of Stage 2 data are fit perfectly by a saturated three-parameter model using GA, GH, and GP.
Figure 4. Comparison of Variance Accounted for by Stage of Decision. Note that Group Ratings explain only a bit more variance than initial ratings for the first four years of the study. In the last year the variance accounted for by Stage 2 increased dramatically. The graph should be read understanding that by Stage 3, all of the variance is explained.

Figure 5. Data, Logistic Model, and Cumulative Cohort as a Function of Stage 1 Rating Category. Applicants have been ranked by their probability of admission based on the logistic model. Acceptance rates by category are graphed surrounded by ±1 Estimated Standard Error. The heavy black line is the admission rate predicted by the logistic regression model. Open squares show the fraction of those eventually accepted accumulated in order of Initial Rating Category.

One can use this graph to compare the relative strength of sets of different ratings. For example, applicants with an initial rating of 2A 1TA, are modeled as having a 40 percent probability of being admitted (thick black line). They have been accepted 42 percent of the time (black square), well within one standard error. The cumulative effect of these ratings is also represented in Figure 5 as the fraction of the final cohort admitted. One can see that 83 percent of the final cohort had initial ratings of 1A 2TA or higher. These ratings were held by only the top 23 percent of candidates. Ratings of 1A 1TA or higher account for 98 percent of those accepted and 38 percent of the applicants.

To save time, an admissions committee may wish to only discuss candidates who after the initial rating stage appear to have a reasonable chance of being admitted. This kind of visual representation aids in setting such thresholds. Applicants with only 2TA and below have a small probability of being accepted into the program, while those with ratings of 1A 3TA or better have a probability greater than 0.50. A reasonable interpretation is to limit discussion to candidates greater than 1A 1TA, approximately 38 percent of the candidate pool.

A logistic model built from the 1992, 1993, 1994, and 1996 admissions data can be used to predict the outcome of the 1997 deliberations. Again the fit is quite good, although some reduction in fit should be expected since we have only one quarter of the actual cases considered. For this set of data, $R^2 = 0.830$, using the Hosmer-Lemeshow goodness-of-fit statistic $\hat{C} = 7.489 (p = 0.914)$. This is evidence that the 3-parameter logistic model is sufficiently robust to predict the proba-
bility (but not the case-by-case result) of admission for each category.

This same technique can be used for modeling the probability of admission by group rating as well (see Figure 6). One can expect the fit of this model to be perfect, since a fully saturated model for four partitions of data can be constructed from the three logistic parameters and a constant. For the first four years of the study, \( \hat{C} = 0.000 \), with a derived Chi-square statistic of \( p = 1.00 \). When the model is fitted to the fifth year of the study, \( \hat{C} = 1.224 \) (\( p = 0.742 \)), indicating a good fit.

**Inter-rater Bias**

Committee members exhibited a range of stringency when compared to the grand average of all raters. We have calculated bias in several ways. In Figure 7, such “average” ratings are graphed at the midpoint of the rating scales by adding together the fraction of students receiving either TD or D ratings. These calculations range from the easiest rater (0.23) to the hardest rater (0.79). The average values for all raters taken together vary little from year to year.

Relevant handicaps and measures of deviation from the mean rating by committee member for each year are reported in Table 4. Cumulative bias, using a comparison of rater mean to grand mean ratings (handicap), shows a \( \pm 1 \) SD spread in rating over the five year study of \( \pm 0.291 \) on a four-point scale. This is small compared to the one point separation on the rating scale. Cumulative bias measured using a handicap\(_{\text{by~case}}\) comparison is reduced further to \( \pm 0.232 \) for \( \pm 1 \) SD. While these values characterize the bias of judges, bias in the admissions process results from the combined effect of four raters’ biases added together. The distribution of individual corrections has a SD average less than .100 of a category. This corresponds to a small change in average rating. The largest accumulated bias for a candidate in each year is listed in Table 4.

**Figure 6**

Comparison of Logistic Model to Group Decision Data


- Model Probability
- Acceptance Rate ±1SE
- Fraction of Final Cohort

**Figure 7**

Distribution of Initial Ratings By Committee Member by Year

- A
- TA
- TD
- D

**Figure 7. Rater Bias.** The distribution of each committee member's rating is shown by initial ratings. Raters are arranged in order of increasing stringency from left to right. Ratios vary in their distribution of assignments of A, TA, TD, and D categories to the candidates. Some raters appear to avoid the high or low end of the rating scale. Although stringency varies considerably, the average number of applicants in each category remains remarkably stable year to year. Raters with fewer than six ratings have been omitted from the graph.
These are also small relative to the four-point scale.

Inter-rater agreement as calculated by the Spearman-Brown reliability formula shows very high agreement between-raters compared to between-candidates with an inter-rater reliability of 0.956.

The more sophisticated Rater-Agreement-Index, which takes into account agreement on a subject-by-subject basis and averages 0.775 over the five years of data, shows a high level of agreement (see Equation 3). This result provides evidence that inter-rater agreement is not a large problem when four raters are used; the random assignment of raters tends to equalize the inter-rater biases. The resulting combination of biases tends to cancel out, leaving little argument for employing a post-hoc correction. Should the number of raters be reduced, such biases would be larger and correction schemes may be warranted.

Discussion

The Admissions Process

The three-stage admissions process is a labor intensive effort which selects few for admission from a large applicant pool. While the number of applicants and committee members has fluctuated over the course of the study, ratings have varied little. This may be because the pool of applicants changes little in strength or it may be a result of the selection process itself. Normally half the committee has served the previous year and their prior experience may have a mitigating influence.

Analysis shows that the initial rating categories can best be thought of as an ordered, but not linear scale. Since we are ultimately trying to accurately predict admission (and not rejection), it makes sense that the Admit category alone predicts a greater fraction of variance than any of the other categories. Committee members have argued for the value of considering the number of positive (A and TA) ratings compared to the number of negative (TD and D) ratings. We find that there is no advantage to preserving the Admit/Deny dichotomy inherent in the initial rating categories; they are statistically interchangeable. A unidimensional scale implies that candidates should not be viewed as having "strengths" and "weaknesses," but simply differing degrees of strength. Just as thermodynamics moved from characterizations of both hot and cold to that of the flow of heat alone, considering candidate strength alone is a more parsimonious model than one of both strengths and weaknesses.

It is useful to divide the candidates into a discrete number of groups using the number of Admit and Tend-to-Admit votes. Applicants with a greater number of these higher ratings have a greater probability of ultimately being accepted. The two lowest rated initial categories, Tend-to-Deny and Deny, appear to have no predictive power for who will eventually be granted admission when the Admit and Tend-to-Admit categories are included in the regression equation.

The four group pool ratings are an intermediate step in the selection process. They are awarded more selectively than initial ratings, with only 10 percent receiving the highest group rating compared to 16 percent for the highest initial rating.

Advantages of Probabilistic Models

Multiple logistic regression equations model the data well, both for initial ratings and group ratings. The regression coefficients represent multiplicative parameters in logistic regression and can be expressed as odds-ratios showing the proportional advantage of one rating over another. These models fit well, even when data from multiple years are aggregated. The model built from the first four years fits the fifth year's data well. This is another indication that the nonlinear logistic equation with only two independent variables is simple, yet robust.

The logistic model provides a framework for comparison of ratings. Figure 5 represents an organized and defensible system for calculating the probability of applicants being accepted into our doctoral program. Based on the fifteen combinations of A and TA ratings ordered by the logistic model, the equation provides a way to assign probabilities even for groups that are sparsely represented in the dataset.\textsuperscript{13}

Of course a calculated probability does not fully determine admission, which requires discussion and individual scrutiny. Yet, it can provide a way to fairly rank the candidates as an aid to further consideration. The extremes are easy to deal with. Certainly four Admits tops the list and four Denys should be at the bottom. Yet to rank the intervening categories, decisions must be made concerning which category is higher than another. It may seem logical to assign an integer value (A=4, TA=3, TD=2, D=1) and average the ratings. This is straightforward, but naive. Deferring to simplicity ignores evidence from past years; these categories are simply not spaced uniformly. In our study, TD and D are indistinguishable; they have the same predictive value. Moreover, the data are better fit by a logistic, or multiplicative set of coefficients than by a linear or additive set. Using an additive integer scheme is clearly inferior, especially when a more representative scheme is available.

Calculating a probability of admission based on initial ratings also serves to allow a fair comparison between candidates at varying positions on the list.

\textsuperscript{13}For example, the category of 3A OTA is formed of only three cases. It could not reasonably continue to maintain a low conditional probability of 0.00 if the study continues. It has an estimated probability of 0.55, based on the regression model.
Errors in logistic models are assumed to follow a binomial statistical distribution (Hosmer and Lemeshow 1989). With a logistic model, one can establish a standard for comparing applicants (for example a probability of ±.05) and assume that combined ratings within such a range are similar (see Figure 5). Comparing applicants from distant positions within the list is also possible; an applicant with 3A 1TA historically has six times the probability of admission than one with 2A 0TA.

The logistic model can also help to group categories so that decisions can be made more easily. For example, applicants with very low ratings should not be discussed unless there is some very compelling reason. The acceptance rate of those with ratings lower than 1A 1TA is small (two candidates in the last five years) even after extensive discussion, yet they represent 58 percent of all applicants. A lower cut-off at this level is wise and fair in that it frees the committee to discuss students who have a more reasonable chance at admission. If any individual member feels a strong need to discuss such a low-ranking applicant, then the case should be made only after the committee has considered all those rated higher.

Another benefit of such models is that they allow for the calculation of the fraction of total variance explained by each step in the admissions process. The initial rating by committee members explains, on average, 39 percent of the variance in acceptance to the program during the years 1992, 1993, 1994, and 1996 (see Table 5). The next stage, the group decision, averages a 21 percent increase. The final stage accounts for 40 percent of the variance since this is where the final decision is made. It is useful to compare this to the amount of labor in each component. Four readers spent a total of approximately 1.3 person-hours/candidate to produce an initial rating (at an average of 20 minutes/rating). The entire committee spends 17 hours to assign group ratings to all the candidates in this period. This multiplies out to 2.1 person-hours/candidate. Final decisions take approximately 4 hours, approximately 0.5 person-hours/candidate.

The largest discrepancy between increased variance and labor content for the first four years of data is in the group rating step, which accounts for 54 percent of the labor content but only contributes 17 percent to the increased variance. One possible interpretation is that these discussions simply place applicants in categories similar to those of the initial ratings, since initial ratings are good predictors of the assigned group ratings. One may argue that the committee is building consensus or community at this stage and that the time is not simply wasted. Yet, this situation should be compared to the result of reducing group ratings to five hours from seventeen hours. In 1997, this time savings was accompanied by a similar increase in variance, interpretable as contributing to the goal of the final decision (Figure 8). The increased variance (17 percent) is close to the total labor content (25 percent) in this stage. In 1997, admissions data from previous years were used as an aid in setting a lower limit on those applicants who would be discussed (at 2TA in 1997). In this way, historical data can be used to increase the efficiency of future deliberations.

Is Inter-rater Bias a Problem?

We are concerned about how important an issue inter-rater bias is to admissions decisions. Fairness can be threatened by the random assignment of raters who systematically differ in stringency. We have looked at these issues quantitatively, describing the variation in mean scores for rater: calculating bias in two different ways and showing that adjusting for rater bias doesn’t change overall reliability of scores. The Spearman-Brown statistic never dips below 0.93. The Rater-

![Figure 8. Comparison of Labor and Variance](image-url)
Table 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>handicap (SD)</td>
<td>0.269</td>
<td>0.346</td>
<td>0.374</td>
<td>0.262</td>
<td>0.203</td>
</tr>
<tr>
<td>handicap by case (SD)</td>
<td>0.181</td>
<td>0.318</td>
<td>0.233</td>
<td>0.230</td>
<td>0.196</td>
</tr>
<tr>
<td>individual correction (SD)</td>
<td>0.108</td>
<td>0.105</td>
<td>0.083</td>
<td>0.074</td>
<td>0.101</td>
</tr>
<tr>
<td>maximum correction to a candidate</td>
<td>0.426</td>
<td>-0.270</td>
<td>0.182</td>
<td>-0.202</td>
<td>0.253</td>
</tr>
<tr>
<td>Speannan-Brown</td>
<td>0.947</td>
<td>0.936</td>
<td>0.962</td>
<td>0.968</td>
<td>0.966</td>
</tr>
<tr>
<td>Speannan-Brown corrected by case</td>
<td>1.000</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
</tr>
<tr>
<td>RAI</td>
<td>0.787</td>
<td>0.770</td>
<td>0.791</td>
<td>0.782</td>
<td>0.746</td>
</tr>
<tr>
<td>RAI corrected by case</td>
<td>0.799</td>
<td>0.783</td>
<td>0.809</td>
<td>0.790</td>
<td>0.752</td>
</tr>
</tbody>
</table>

Table 4. Measures of Inter-Rater Agreement. All measures of inter-rater agreement are high. One Standard Deviation about the mean for each of the five years of the study have been calculated. Calculations were performed with a value of 1.00 separating each initial rating category. Handicap by total compares the mean ratings for each committee member about the mean for all raters. Handicap by case compares the mean discrepancy for each case rated by a rater with the three other readers. This is a more accurate estimate of agreement with others and is consistently smaller. Individual correction is derived by combining the 4 raters' handicap by case to each candidate. This is an estimate of how these handicaps by case impact the candidate's rating. With a standard deviation of ± 0.094 over five years, post-facto correction of ratings would have little impact.

Table 5

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of candidates</td>
<td>139</td>
<td>107</td>
<td>95</td>
<td>111</td>
<td>113</td>
<td>140</td>
</tr>
<tr>
<td>committee members</td>
<td>12</td>
<td>16</td>
<td>13</td>
<td>16</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td># of Initial ratings</td>
<td>558</td>
<td>429</td>
<td>378</td>
<td>447</td>
<td>453</td>
<td>540</td>
</tr>
<tr>
<td>Initial Rating hours/member</td>
<td>16</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Group meeting hours</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>19</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Final Decision hours</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>total in hours/member</td>
<td>36</td>
<td>28</td>
<td>28</td>
<td>33</td>
<td>31</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Initial Ratings</td>
<td>1.34</td>
<td>1.34</td>
<td>1.33</td>
<td>1.34</td>
<td>1.34</td>
<td>1.29</td>
</tr>
<tr>
<td>Stage 2: Group Ratings</td>
<td>1.47</td>
<td>2.24</td>
<td>2.05</td>
<td>2.74</td>
<td>2.13</td>
<td>0.54</td>
</tr>
<tr>
<td>Stage 3: Final Rating</td>
<td>0.26</td>
<td>0.60</td>
<td>0.41</td>
<td>0.72</td>
<td>0.50</td>
<td>0.32</td>
</tr>
<tr>
<td>total in person-hours/candidate</td>
<td>3.06</td>
<td>4.18</td>
<td>3.79</td>
<td>4.80</td>
<td>3.96</td>
<td>2.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Initial Ratings</td>
<td>0.44</td>
<td>0.32</td>
<td>0.35</td>
<td>0.28</td>
<td>0.34</td>
<td>0.60</td>
</tr>
<tr>
<td>Stage 2: Group Ratings</td>
<td>0.48</td>
<td>0.54</td>
<td>0.54</td>
<td>0.57</td>
<td>0.54</td>
<td>0.25</td>
</tr>
<tr>
<td>Stage 3: Final Rating</td>
<td>0.08</td>
<td>0.14</td>
<td>0.11</td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: Initial Ratings</td>
<td>0.395</td>
<td>0.423</td>
<td>0.420</td>
<td>0.334</td>
<td>0.393</td>
<td>0.410</td>
</tr>
<tr>
<td>Stage 2: Group Ratings</td>
<td>0.178</td>
<td>0.114</td>
<td>0.059</td>
<td>0.222</td>
<td>0.143</td>
<td>0.418</td>
</tr>
<tr>
<td>Stage 3: Final Rating</td>
<td>0.427</td>
<td>0.463</td>
<td>0.521</td>
<td>0.444</td>
<td>0.464</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Table 5. Labor Breakdown of the Admissions Process. The workload of admission committee members is substantial, averaging 31 hours/member for the first four years of the study and 20 hours for the last. Candidates get the benefit of many person-hours of consideration in addition to the efforts of the admissions office staff. Of the three stages of admission, Stage 2 accounts for the smallest increase in variance for the first five years, leaving a large fraction to be contributed in the last few hours of meeting time. Changes in the admissions process reduced the labor required to consider each candidate from 3.96 person-hours during the first four years to 2.14 person-hours in 1997.
Agreement-Index of 0.75 is also quite high.

In what ways can post-facto corrections for stringency be performed? What is their effect? Subtracting rater handicap from all initial ratings would have no effect on categorization, since handicaps do not exceed 0.50 for any rater. Applying all rater handicaps to a candidate makes the most sense. The RAI increases by a value of 0.01 when accounting for stringency in this way (see Table 4). These measures indicate that variability in ratings by raters is relatively small, and that overall bias has little impact on the eventual decisions.

However, raters often experience themselves as having very different views of a candidate than do their colleagues. Many feel as though they’re arguing for candidates where others have missed obvious strengths or are raising critical doubts about candidates whom others find stellar. This manifests itself in a desire to discuss such candidates anew in the group rating stage. How can we reconcile these reassuresing overall quantitative measures of rater agreement with the more conflicted experiences of raters? Can we design an admissions process to make use of these two types of evidence—summarized quantitative and particular experiences? To do this, we need to better understand the experiences of raters on the committee. How do committee members see their ratings incorporated into the pooled ratings of candidates and the eventual admission decisions?

The reassurance of a high statistical agreement index is of little consolation when others question your judgment compared to theirs. How much do such ratings disagree? In fact, 63 percent of the time ratings differ by no more than a category for any candidate. With our coarse, four-category scale, there should be such high agreement. Sorting a continuum into four ordinal categories will always split votes when people are between categories.14

Several other sources for rater disagreement can be imagined. Files early in the rating process, especially by first time committee members, can be read with a different eye than those later in the process.15 Some raters do not use the entire four-category scale, tending to avoid either the highest or lowest categories (see Figure 7). Such “scale shrinking” can occur when less confident members of the committee may find it easier to err on the side of lower ratings (Rudner 1992). Our policy of asking the most positive rater to take the lead when presenting a candidate may motivate some to reduce ratings rather than to serve as a candidate’s primary advocate. This effect can become more pronounced after the “calibration meeting” as new committee members come to understand the advocacy role entailed by a high rating. Some members feel that giving a low rating to a candidate is objectionable. This tends to compress their ratings at the high end. These tendencies toward scale shrinking might be alleviated by explaining to raters that not making use of the full range of ratings reduces their contribution to the decision process. Raters should use the full scale since the job of the committee is ultimately to make these important distinctions between candidates. If members do not contribute to making such distinctions early in the process, these decisions must be made at a later stage.

Our sparse, four-category rating scale tends to reduce the impact of rater bias. For a scale with more categories, such as a ten-point scale used by some programs, correction would make more sense. Corrections to ratings appear to be unwarranted in our application process, yet attention could be paid to lowering the disparity between the average ratings of committee members. A close monitoring and sharing of handicaps throughout the admissions process may serve to reduce this difference. If the number of raters per folder was reduced to below four, inter-rater agreement could become a more serious problem and post-facto adjustments might increase the fairness of decisions made from these ratings. Not correcting for such disparities would leave ratings to the chance biases of those who rated the folder (Cason and Cason 1989). Correcting for rater bias could also allow for fewer readers while maintaining the reliability of the process.

Conclusions

Our five-year study has unclouded some of the mystery shrouding our doctoral admissions process, contributing in small part to an understanding of a huge industry with an estimated 2,263,000 undergraduates and 431,000 graduate students entering school each year (American Almanac 1993). College and university admissions processes, methods, and policies are labor intensive and not widely publicized. Such secrecy probably serves to limit the identification of common problems, because so little is published about how admissions decisions are made. Freely sharing practices and innovations could increase reliability; fairness may be inhibited by such secrecy.

The collection and analysis of our data reveals patterns that appear to remain stable over time in our three-stage process. The four-category, initial rating by four independent readers produces similar distributions each year. The combined raw scores of candidates can be

---

14 If an applicant has qualities that rank him/her between Admit and Tend-to-Admit, we should expect a split vote and ratings of 2A and 2TA. Larger discrepancies among raters require more explanation.

15 A careful search for such a trend, either positive or negative, has not been found with regard to when in the admissions process a candidate was rated.
used to predict a reasonable cut-off for selection of candidates who should be considered in the next stage. A two-variable logistic regression equation models such data well and has certain advantages over using raw scores alone. The model allows comparison of rating categories by odds-ratio and can consistently calculate the probability of admission even in cases where historical data are sparse. The model can be used to rank candidates without making arbitrary assignments concerning the equivalence of rating categories.

A logistic model can also be used to account for the proportion of variance explained by each successive stage in the admissions process. Roughly 40 percent of the variance in the final selected cohort is explained by the initial ratings. The group rating process typically appears to be the weakest step, explaining only one-fifth of the final variance; yet, it can consume more than half the labor content of the admissions process. Reducing the number of candidates discussed in this second stage of admissions appears to make good use of prior data and of committee members' time.

It may be that group discussions satisfy a desire by committee members to help justify their high ratings or to share the burden of rejecting candidates. It is troubling to give low ratings to candidates who clearly have worked hard preparing an application and are seeking admission. Committee members may also feel a need to make an independent decision based on their own assessment of the candidate and may not be willing to accept others' ratings as proxies for their own. Measures of inter-rater agreement show that members can well rely on the judgments of others. Agreement is high and rater bias in stringency tends to average out over four initial readers. Reconsidering the reasoned ratings of readers is not a good use of time.

We have looked at admissions in one graduate program over a five-year period. Other programs may show results different from ours. Yet, the tools developed here for analysis may be very useful for others to apply. We recommend that all admissions programs that use subjective ratings examine their decisions using quantitative tools. For programs that categorize or quantify qualities of applicants and use such data in discussions and multistage deliberations, these steps will serve to increase fairness and efficiency:

- Care should be taken to preserve records of intermediate classifications and decisions for later analysis. Such records need not preserve the identity of raters or candidates.
- Categorizations and ratings should be analyzed on the basis of final decisions.
- Logistic models can be developed which gauge comparable merit of rating categories, to establish thresholds for consideration, and to rank candidates based on the historical probability of admission.
- Rater bias should be calculated and its impact on decisions explored fully. Should it be found to play a substantive role, rater training, careful monitoring to reduce bias, or post-facto corrections can be used to reduce systematic error and increase fairness for candidates.

Applying for admission to a college or university is a significant and important step in a person's life. Compressing one's experience and dreams onto paper and mailing it off to a murky fate is vexing. Admissions officers and committee members attempt to select the best candidates possible, often from very large pools. Utilizing quantitative tools does not necessarily reduce the humanity of the process. It may serve to increase fairness and help by concentrating deliberations on those who are best matched to the strengths of the program.

Acknowledgments

The authors wish to thank several of their colleagues for help with this study. Data were prepared and cross checked by Jan Still, Helen Spencer, and Judith Peritz. Library research was carried out by Joelle Pelletier. Appropriate statistical tests were suggested by Judith Singer and Robert Tai. Ideas and feedback from our colleagues Eleanor Duckworth, Catherine Elgin, Marcus Leiberman, Donald Oliver, Roland Spence, Terrence Tivnan, Dean Whitta, David Perkins, and Cecilia Wu have aided us in addressing issues more clearly.

References


Hagedorn, L. S. 1996. Rethinking admissions criteria in graduate and professional programs. *New Directions for Institutional Research* 92: 31-44.

Hall, F. R. 1992. Correlating students' undergraduate science GPAs, their MCAT scores, and the academic caliber of their undergraduate colleges with their first-year academic performance across five classes at Dartmouth Medical School. *Academic Medicine*. 67(2): 121-123.


A Study of Canadian and American Admissions Equity at Brigham Young University

Jeffery M. Tanner and R. Wayne Shute

Introduction

For years, parents, students, and educators from Canada have felt that Canadian applicants to Brigham Young University (BYU) have been required to qualify for admission at a higher standard than their U.S. counterparts. BYU requires the American College Test (ACT) as part of the application process. Canadians have expressed concern particularly about the Social Science area of the ACT because the material deals primarily with the history of the United States, not Canada. In addition, many Canadians have felt that the grading practices in Canada were more rigorous than those in the United States. Canadian parents, students, and educators also have felt that equating like grades between the United States and Canada discriminated against Canadian students.

Statement of Purpose

The purpose of this study was two-fold: 1) to compare the ACT examination results of Canadian students with their American counterparts to see if the ACT is biased against Canadian students; and 2) to compare Canadian and American students' academic achievement while the students are in their first year at Brigham Young University as a way of determining if high school grades of Canadian students were being undervalued in the admissions process at BYU.

Educational Equity Between the United States and Canada

Hurn and Burn (1982) indicated that comparing the outcomes of different educational systems is like comparing apples and oranges. They go on to explain that there are two basic reasons why comparisons are problematic at best. First, there is the problem of educational objectives not being weighted the same in different societies. Second, there is the problem of selectivity of those things that will be compared. Even with the difficulties mentioned, there have been attempts to compare, at least on the periphery, the educational systems in the United States and Canada.

When looking at similarities in education between Canada and the United States, King (1982) found that American students who had enrolled at unidentified universities in Ontario, Canada, did as well as, or better than, students from Canadian provinces. Where admission standards were somewhat lower for American students, the students tended to achieve at a lower level than did students from Canadian provinces. The better American students were quite competitive with university entrants from Canada.

Redd and Riddle (1988) prepared a report for the U.S. Congress in 1988 entitled Comparative Education: Statistics on Education in the United States and Selected Foreign Nations. Among other things it was shown that:

1. On a proportional basis, similar expenditures are made for education at all levels.
2. Enrollment rates for both countries for primary, secondary, and postsecondary students are similar.
3. There are similarities when looking at the percent of postsecondary students by field of study.
4. The lengths of the school year for elementary and secondary education in the two countries are within two weeks of each other.
5. Achievement scores in science and mathematics, when comparing 13-year olds and 17-year olds, are very similar.

With all of these similarities, one might assume that there would also be a similarity in the performance of university students with similar grades and test scores from these two countries. If not, then the comparability of grades and test
scores between the two countries would be suspect. In addition, such results would add data to the question of whether or not the Canadian system of education better prepares a student for a university education than does the American system.

**Method**

The method of research for this study was comparative using quantitative data. It employed a matched pairs design to determine if there is bias in the BYU admissions selection process against Canadian applicants. Each student who attended high school in Canada and who met the selection criteria was matched against a student whose high school education was in the United States.

The initial selection of the population used in this study consisted of all new freshmen students who attended Brigham Young University for at least two consecutive semesters and who began their enrollment after spring term 1987 and prior to summer term 1991. The students selected were citizens of Canada or the United States at the time of their initial enrollment at BYU. The records of all Canadian students who met these criteria were used. The record of each Canadian student was matched with the record of a student from the United States who also met the initial selection criteria.

Match elements included age, gender, the college of the students' initial major, high school grade point average, university entry year and term, ACT Composite scores, and that the students took the same version of the ACT (old, prior to 1989 or new, 1989 or later).

The Brigham Young University first year cumulative grade point average was the criterion used to determine whether or not there was bias in the BYU admissions process for Canadian students. It was recognized that grade point average is not determined in an absolute way and that subjectivity, can be, and often is, a part of all grading practices. However, it was assumed that the broad range of course sections taken minimized the effects of individual grading bias so as not to be a significant factor.

The assumption was that if the admissions selection criteria being used by the BYU Admissions Committee incorrectly under-valued the academic preparation of Canadian students, that under-valuing would be demonstrated by their superior first year university grade performance when compared to the first year university grade performance of U.S. students with whom they were matched.

The mean scores of all Canadian students who had taken each section of the old ACT were compared with their matched counterpart from the United States who had also taken the old ACT. The mean scores for the two groups were subtracted and the difference of the Canadian students' scores compared with their American counterparts is shown in Table 2. The same process was followed for those Canadian and American students who had taken the new ACT.

Table 3 provides data regarding the differences in mean ACT standard scores looking at three different variables. The statistical inferences noted in the data come from performing a MANOVA analysis on the three different variables. The first variable is the type of ACT examination taken, old or new. These results are shown under the column titled Type. The second variable is the combining of all ACT examination scores, by examination section, for Canadian students, and then comparing their performance with all United States students whose scores had been combined in the same way. These results are shown under the column titled Group. In the final instance, performance of the Canadian students who had taken the old ACT examination were compared with their American counterparts who had taken the old ACT examination. Then the same thing was done for the Canadian and American students who had taken the new ACT examination. These results are shown under the column titled Type and Group.

Table 4 shows the first year grade performance of Canadian students versus their American counterparts at BYU.

**Findings**

A paired comparison t-test was performed on all match items except those that were exact (i.e., gender, semester and year of initial BYU enrollment, college of first major). The purpose was to see if like individuals from Canada and the United States had been selected as part of the match process. Table 1 illustrates the outcome of that test.

The first variable is the type and group. The second variable is the combining of all ACT examination scores, by examination section, for Canadian students, and then comparing their performance with all United States students whose scores had been combined in the same way. These results are shown under the column titled Type. The second variable is the combining of all ACT examination scores, by examination section, for Canadian students, and then comparing their performance with all United States students whose scores had been combined in the same way. These results are shown under the column titled Group. The purpose was to see if like individuals from Canada and the United States had been selected as part of the match process. Table 1 illustrates the outcome of that test.

The mean scores for all Canadian students who had taken each section of the old ACT were compared with their matched counterpart from the United States who had also taken the old ACT. The mean scores for the two groups were subtracted and the difference of the Canadian students' scores compared with their American counterparts is shown in Table 2. The same process was followed for those Canadian and American students who had taken the new ACT.

Table 3 provides data regarding the differences in mean ACT standard scores looking at three different variables. The statistical inferences noted in the data come from performing a MANOVA analysis on the three different variables. The first variable is the type of ACT examination taken, old or new. These results are shown under the column titled Type. The second variable is the combining of all ACT examination scores, by examination section, for Canadian students, and then comparing their performance with all United States students whose scores had been combined in the same way. These results are shown under the column titled Group. In the final instance, performance of the Canadian students who had taken the old ACT examination were compared with their American counterparts who had taken the old ACT examination. Then the same thing was done for the Canadian and American students who had taken the new ACT examination. These results are shown under the column titled Type and Group.

Table 4 shows the first year grade performance of Canadian students versus their American counterparts at BYU.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample Size</th>
<th>Canadian Students</th>
<th>American Students</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGPA</td>
<td>277</td>
<td>3.36</td>
<td>3.36</td>
<td>0.00</td>
</tr>
<tr>
<td>ACT Comp.</td>
<td>277</td>
<td>24.55</td>
<td>24.34</td>
<td>0.21</td>
</tr>
<tr>
<td>Age</td>
<td>277</td>
<td>18.00</td>
<td>18.02</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
Table 2: Differences in ACT Mean Scores (Canadian vs. U.S. Students)

<table>
<thead>
<tr>
<th>Test Section</th>
<th>Canadian Student Old ACT Mean Score Difference</th>
<th>Canadian Student New ACT Mean Score Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>0.0</td>
<td>-0.3</td>
</tr>
<tr>
<td>Mathematics</td>
<td>+0.3</td>
<td>-1.0</td>
</tr>
<tr>
<td>Social Science/Reading Comp</td>
<td>-0.7</td>
<td>+0.3</td>
</tr>
<tr>
<td>Natural Science/Science Reasoning</td>
<td>+1.9*</td>
<td>0.0</td>
</tr>
<tr>
<td>Composite using all sub-scores</td>
<td>+0.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>Composite score without Social Sci./Reading Comp</td>
<td>+0.7</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

*p < .001

Table 3: ACT Comparisons

<table>
<thead>
<tr>
<th>ACT</th>
<th>Type</th>
<th>Group</th>
<th>Type and Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>-3.6**</td>
<td>-0.2</td>
<td>-0.3</td>
</tr>
<tr>
<td>Test 2</td>
<td>-1.5*</td>
<td>-0.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>Test 3</td>
<td>-5.5**</td>
<td>-0.2</td>
<td>+1.0</td>
</tr>
<tr>
<td>Test 4</td>
<td>+0.9</td>
<td>+1.0*</td>
<td>-1.9*</td>
</tr>
<tr>
<td>Test 5</td>
<td>-2.5**</td>
<td>+0.1</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

Note: Test 1 = old English vs. new English. Test 2 = old Math vs. new Math. Test 3 = old Social Science vs. new Reading Comprehension. Test 4 = old Natural Science vs. new Science Reasoning. Test 5 = old Composite vs. new Composite.

*p < .05. **p < .001

Table 4: Mean BYU Grade Performance Comparison

<table>
<thead>
<tr>
<th>MEANS</th>
<th>Sample Size</th>
<th>Canadian Students</th>
<th>American Students</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYU-GPA</td>
<td>277</td>
<td>2.96</td>
<td>2.83</td>
<td>+0.13</td>
</tr>
</tbody>
</table>

Table 5: Paired Comparison t-test

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Mean difference</th>
<th>Standard error of mean</th>
<th>t-value</th>
<th>Probability Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>BYU-GPA</td>
<td>0.126642</td>
<td>0.051779</td>
<td>2.44</td>
<td>0.0151</td>
</tr>
</tbody>
</table>

Table 5 provides the data relating to a paired comparison t-test on the Brigham Young University grade performance of Canadian students when compared to their matched counterpart from the United States.

A review of the information and tables in this findings section reveals that:

1. There was an excellent match between Canadian and American students selected for this study (Table 1).
2. There was no indication in the data that either the old or the new ACT examination discriminated against Canadian students (Table 2).
3. No significant difference in performance was determined to exist between Canadian and American students on either the individual test areas or the composite score on the new ACT (Table 2).
4. Statistical significance was demonstrated showing that the old ACT and the new ACT examinations are different (Table 3).
5. When combining old and new ACT examination results for Canadian students and comparing them to the combined old and new ACT examination results for American students, only in the Natural Science/Science Reasoning areas is there a statistically significant difference in performance demonstrated favoring students from Canada (Table 3).
6. When comparing the type of ACT examination (old or new) taken by Canadian students with the type of ACT examination taken by American students, only in the Natural Science/Science Reasoning areas is a statistically significant difference in performance demonstrated favoring students from the United States (Table 3).
7. When comparing first year grade performance at Brigham Young University, Canadian students performed at a statistically higher level than U.S. students (Tables 4 and 5).

Discussion

It is clear from the data gathered and evaluated that the old ACT examination did not discriminate against students from Canada in the area of Social Science. There was no statistically significant difference between the performance of Canadian students when the results of that section were compared with the results achieved by students from the United States. Nor was there any major difference when the ACT composite scores were generated using all sub-scores or using all sub-scores minus the Social Science score.

Since the old ACT did not discriminate against Canadian students, the new ACT could not correct an old ACT deficiency that did not exist. The concerns that had been expressed by Canadian parents, students, and educators were shown to have been more imagined than real. In addition, the new ACT was not shown to have introduced bias against Canadian students in any of its sections.

When evaluating the information regarding the BYU first year grade performance of Canadian students and their matched counterparts from the United States, there is a statistically significant difference in the performance of the two groups. The performance of the Canadian students at 2.96 as compared to their matched counterparts from the United States at 2.83 was shown to be significant at the 0.015 level. The null hypothesis could therefore be rejected at the 0.015 level based on the data gathered and evaluated.

This study has provided evidence that Canadian students do outperform their American counterparts at least during the freshman year. A reasonable person would probably agree with Canadian parents and educators that a “B” grade from Canada is somewhat higher than a “B” grade from the United States when looking at outcomes as measured by first year college grade performance. There would probably be concurrence with the claims of Canadian parents and educators that adjustments should be considered for the admissions selection process at BYU to account for the higher performance of Canadian students.

The reasons for the improved Canadian students’ performance were not part of this study. However, some possibilities do come to mind. It may be that the improved performance was the result of a better educational approach in the secondary education of Canadian students. It may be that Canadian students, feeling a need to be exemplary university citizens and represent their country well, are motivated to study harder and longer in order to do better. It may be that LDS students from Canada are a unique subset of the Canadian population as a whole and/or a unique subset of the LDS student population as a whole and their performance is the result of that uniqueness rather than any experience they may have had in their secondary educational experience. The somewhat higher performance levels may be a result of any or all of these factors or may have nothing to do with them at all. Nevertheless, the Canadian students in this study did outperform their American counterparts.

There are two related items that were demonstrated in the statistical evaluation. First, it was shown that there is a very high degree of probability that additional statistical studies would produce results similar to those demonstrated in this report. Second, the difference between a 2.83 and a 2.96 grade point average is statistically validated because the difference would continue to appear if the same research procedures were followed a second or third time.

The thing that the statistical research does not demonstrate is how important the difference is, or what value should be attributed to the difference that was found. This is best illustrated in the following: “Whether or not the magnitude of the difference between Mu of A and Mu of B is of any practical importance is a question that cannot be answered by the statistical test. This is a question that only the researcher can answer after consideration of non-statistical information” (Thompson 1996).

Most colleges and universities would consider significant performance differences being achieved at about 35 hundredths of a grade. This is the difference between a “B” and a “B+” or a “C-” and a “C.” Grade keys printed on college transcripts deal with these levels of difference. In addition, most students use letter grades when describing their own performance. It is more common for them to describe themselves as “B-” students than to say that they are “2.8” or “2.9” students.

It was important to find statistical evidence that there is a difference in grade performance between students from the United States and Canada, and that Canadian parents and educators were correct in assuming that such a difference existed. At the same time, there is also the realization that the relative smallness of the effect size (2.83 to 2.96), is not important enough to warrant major changes in the admissions procedures at Brigham Young University. Instead there is the feeling that by any practical standard the Brigham Young University Admissions Office is about as equitable, in admissions considerations, as is practical.†
References


C.E.I. Education Specialists, Inc.

International Transcript Evaluations by Professors

Federal Court Admitted Experts in International Education
Prompt Service • 30 Years of Experience
Reasonable Cost • Member of AACRAO, NAFSA, IIEEA

• All disciplines
• Countries — course by course, subject by subject
• Degree authentication

Evaluations accepted by federal and state agencies/licensing boards, colleges, universities, and industry

For more information, contact
10 Legend Lane, Houston, TX 77024-6003 or
P.O. Box 19756, Houston, TX 77224-9756
Tel. 713-952-6753; Fax 713-467-2474
E-mail: sant2@aol.com

29

Spring 1999
Educational Consortia — A Longitudinal Study

Marybelle C. Keim, Ph.D.

Multi-purpose consortia in higher education are defined as voluntary, formal organizations of two or more member institutions that have more than one program, are administered by a professional director, and receive continuing membership support (Connick and LaRocco 1983). Franklin Patterson (1974) attributed the origin of the movement to the Claremont Colleges in 1925 and to the Atlanta University Center in 1929. The growth of consortia was slow until the late 1940s and the 1960s led to a rapid increase in numbers (Neal 1988). Mathews (1974) posited that consortia "grow best out of the internal interest of faculty, staff, and students" and that "consortia are concerned with growth and innovation." O'Neil (1974) suggested that many consortia began with projects, in the hope of reducing costs and expanding the educational opportunities for students.

According to Grupe (1975), a sequence of phases is necessary to establish cooperative arrangements. The phases are (1) exploration, (2) planning, and (3) implementation. In the first phase, a single committed individual who has a sense of direction is needed to initiate the effort. There must be potential institutional clientele, in geographic proximity, with a common concern. In the second phase, there should be a planning committee of presidents and other representatives, institutional support, and approval of the planning committee’s recommendations. The third phase includes the selection of an executive officer, suitable quarters, formal activation of the consortium, and initiation of the first programs by the executive director.

Among the benefits of consortium membership are:

- decreased redundancy of expertise;
- enhanced quality and breadth of service offerings, often without added costs;
- a broader range of options in responding to change;
- continuity during times of staff turnover on campus;
- a different kind of reallocation process;
- sharing of overhead costs;
- sustained efforts during peaks and valleys;
- opportunities to build skills, keeping staff refreshed and engaged in their work;
- more rapid diffusion of best practices across campuses;
- the power of combined purchasing volumes;
- reduced administrative costs; and
- an enhanced sense of professional community within higher education (Shafer and Reed 1996).

Numerous articles, books, dissertations, reports, and conference papers have been published about consortia and interinstitutional cooperation among universities, colleges, community colleges, and public schools since the early 1950s. Lewis Patterson contributed significantly to the movement, as a frequent author, as editor of The Acquaintain—a newsletter about consortia, as well as a consortium executive at the Kansas City Regional Council for Higher Education (KCRCHE) and at the American Association for Higher Education (AAHE). In an early ERIC research report, Patterson (1970) detailed the history of consortia, stages of development, purposes, various types, and funding requirements. He (1971) subsequently wrote his dissertation about governance in selected cooperative arrangements and an additional AAHE/ERIC report (1979) on the costs and benefits of interinstitutional cooperation. Other early proponents of consortia were Grupe (1975) (although he was chagrined that "consortia do not have a good reputation"); and Neal (1984), who in a 1982 informational presentation for the Society for College and University Planning, explained the potential of consortia to the conference attendees. Neal (1988) also edited a book, Consortia and Interinstitutional Cooperation, which detailed the functions and roles of consortia.

Marybelle C. Keim is a Professor of Higher Education in the Department of Educational Administration and Higher Education at Southern Illinois University at Carbondale. She earned a Ph.D. in Administration and Higher Education at Michigan State University and an M.A. in Educational Psychology at the University of Northern Colorado. Her interest in Consortia began in the early 1980s when she was employed by the Kansas City Regional Council for Higher Education as an Institutional Researcher and Director of the College Information Service. She has also served as a Board Member for the Association for Consortium Leadership.
Few large scale studies have been conducted about consortia. Lewis Patterson (1971) studied 51 consortia using their articles of incorporation, bylaws, and statements of agreement as his resource documents. Franklin Patterson (1974) based his book, *Colleges in Consort: Institutional Cooperation Through Consortia*, on field visits to 26 consortia and on documents, correspondence with consortium directors, graduate theses, conference reports, and newsletters. The basis of Ryan’s (1981) remarks at the annual meeting of the American Council on Education (ACE) in 1979 were 14 consortia that he worked with at Indiana University.

Much of the research has involved case studies, usually dealing with a limited number of consortia. Examples include Shafer and Reed (1996) who presented sketches about the Associated Colleges of the South, Five Colleges, Claremont Colleges, and the Boston/Mellon Consortium; Elkin (1982) who described the Great Lakes Colleges Association; Hanson (1986) who studied the Associated Colleges of Central Kansas; Poland (1986) who looked at one Virginia Consortium for Continuing Higher Education; Belovarac (1984) who investigated the faculty and administrative roles in the development of the Erie Consortium of Colleges; Breuder (1996) who discussed a merger that became the Pennsylvania College of Technology; and Offerman (1985) who determined the factors leading to the termination of three higher education consortia.

Specialized kinds of consortia have also been discussed in the literature, including library/learning resource centers (Voegel 1986; Woodworth 1989), research (Holland 1990), technology (DeLoughry 1994), assessment (Astin and Ayala 1987), international agendas (Fifield and Sakamoto 1987), contracts (Ferris 1991), telecourse production (Zigerell 1991), nursing (Algren and Hockenberger 1986), adult learners (“Partnerships” 1989), school-college partnerships (Wilbur 1985), and university-industry partnerships (Ferris 1991). Thought pieces have been written (Groves and Groves 1983), and literature reviews have been conducted (Christensen and Wylie 1991; Konrad and Small 1986). Despite many publications, not one gives a current global perspective of consortia in higher education; this article was written to fill that void.

**Method**

The purpose of this research was to provide a longitudinal study of consortia during the past 15 years. To accomplish this purpose, data contained in the 1983 *Consortium Directory* (Connick and LaRocco 1983), the 1991 *Consortium Directory* (Love and Barnett 1991), and the 1996 *Consortium Directory* (Dotolo and Beltz 1996), published by the Association for Consortium Leadership (ACL), formerly the Council for Interinstitutional Leadership (CIL), were analyzed.

The 1983 *Directory* contained information about 134 consortia located in 37 states, the District of Columbia, Canada, Costa Rica, and Guam; the 1991 *Directory* listed 121 consortia located in 40 states, the District of Columbia, Australia, and Canada; and the 1996 *Directory* included data about 80 consortia located in 31 states, the District of Columbia, and Canada.

The data were summarized using a computer database; frequencies, percentages, means, and medians were compiled where appropriate. The results and conclusions, along with the corroborating data, are presented under these subheadings: characteristics of consortia, directors, activities, and comparisons with previous research. It must be noted that no attempt was made to verify the data contained in the *Directories* and that the data were used in this analysis as if they were accurate. While the data collection method used in compiling the *Directories* was not explained, it was assumed that questionnaires had been mailed to each of the consortium directors and that the directors had provided the information printed in the documents.

**Results and Conclusions**

**Characteristics of Consortia**

The total number of consortia declined between 1983 and 1996, but the average number of members in each consortium increased.

The number of institutions in each consortium ranged from 2 to 345 in 1983, from 2 to 1,500 in 1991, and from 3 to 1,500 in 1996. The mean number of members in the 1980s was 22 and increased to 46 in 1991 and declined slightly to 45 in 1996. When the largest consortium was not included in the calculations for 1996, the mean was 26 institutions per consortium and the median was 13.

A number of consortia disappeared between 1983 and 1996.

Sixty-six consortia appeared in all three editions of the *Directories*. Of the survivors, 36 (54 percent) gained members, with a range of 1 to 313, a mean of 15, and a median of 2. Five (8 percent), lost members, with a range of 1 to 230, a mean of 47, and a median of 1. The numbers of members did not vary for 25 (38 percent). Eleven consortia (13 percent) had name changes between 1983 and 1991 and 14 (21 percent) changed names between 1991 and 1996.

The word, consortium, appears in few program titles.

In 1983, 50 out of 134 consortia (37 percent) had the descriptor “consortium” in their titles. By 1991, 38 out of 121 (31 percent) used such a designation, and by 1996, 24 out of 80 (30 percent) included the descriptor.
Consortia have been in existence for several decades, but most were founded in the 1960s and 1970s.

The oldest consortium was reported to have begun in 1867 and the most recent was initiated in 1992. One began in 1896; one started in the 1910s; 2 had founding dates in the 1920s; 2 had 1930s dates; 4 began in the 1950s; 28 in the 1960s; 23 in the 1970s and 11 in the 1980s. Altogether, 64 percent of consortia were founded between 1961 and 1980. However, there were several discrepancies in the founding dates in the Directories. In fact, of those that could be matched between 1983 and 1991, 27 or 32 percent had different years of origin, with a range of 1 to 16 years difference, and a mean of nearly 4 years. Between 1991 and 1996, 8 had differing dates ranging from 1 to 7 years, with a mean discrepancy of 3.5 years. The differences may have been caused by inaccurate reporting or poor record keeping.

Consortia are most frequently located in eastern and midwestern states.

The state with the most consortia was Ohio (n=8), followed by Massachusetts (n=7), Illinois (n=6), Pennsylvania (n=6), and Virginia (n=5). Having 4 consortia each were California and the District of Columbia. Four states each had 3 consortia; 6 states each had 2 consortia; and 16 states and countries each had one consortium.

Most consortia are governed by a Board of Directors.

This information was not included in the 1983 edition, but from the later documents it was determined that in 1991 nearly 40 percent (n=48) were governed by a Board of Directors. Twenty-eight percent (n=34) did not furnish the data; 13 percent (n=16) had a Board of Trustees; 4 had an Executive Committee; and the other consortia used 13 additional governing arrangements.

In 1996 55 percent (n=42) had a Board of Directors, 14 percent (n=11) a Board of Trustees, 6 a Board, 4 did not furnish the data, 3 a Council, 2 an executive committee, and the remaining 12 consortia used other governance organizations.

Consortia are funded in a number of ways.

These data were contained only in the 1991 and 1996 Directories. In 1991, 17 consortia had one source of funding, relying solely on dues for their support. Twenty-seven consortia reported two sources of funding, using dues and grants, or dues and projects, or dues and fees, or other combinations. Twenty-four consortia listed 3 sources with 20 different arrangements. Six consortia listed 4 and 5 sources of funding. Of the funding data furnished for 90 consortia, nearly 37 percent, (n=33) were supported in part by corporate/foundation grants, or by federal grants. Most consortia executives did not furnish the percentages of each of their funding sources, but for those who did, the level of external funding ranged from 1 percent to 85 percent of their total budgets.

In 1996, 17 consortia relied completely on dues as their source of funding. Twenty-three used two sources of funding, with a variety of combinations—dues and services, dues and grants, state appropriations and grants, state appropriations and grants, and other, and contributions and other. Sixteen listed three sources, 10 had 4 sources, and 4 were supported by 5 sources.

The 1996 document listed funding sources by percentages. In the dues category, the mean was approximately 52 percent and the median was 60 percent. In other words, consortia, on average, relied on dues for 50 to 60 percent of their budgets. Eleven were not supported in any way by dues. In the state appropriations category only 6 received state monies; the range was from 10 to 100 percent with a mean of 62 percent and a median of 82 percent. In the grants category, 35 received grants, with a range of 1 percent to 94 percent, a mean of 39 percent, and a median of 26 percent. In the contributions category were 16 consortia, with a range of 4 percent to 80 percent, a mean of 32 percent, and a median of 11 percent. In the services category were 26 consortia, with a range of 1 percent to 94 percent, a mean of 30 percent, and a median of 18 percent. In the “other” category were 32 consortia, with a range of 1 percent to 60 percent, a mean of 15 percent, and a median of 6 percent.

Consortia budgets vary a great deal.

This information was included in the most recent Directories and was furnished by 90 consortia in 1991 and by 69 in 1996. In 1991 the range was from $8,800 to $20 million, with a mean of approximately $1,100,000, and a median of $200,000. When the 4 consortia with budgets of more than $9.5 million each were not included in the calculations, the mean dropped to $485,700.

In 1996 the range was from $4,000 to $30 million, with a mean of $1,600,044, and a median of $287,237. When the 2 consortia with budgets each in excess of $25 million were excluded from calculations, the mean was $673,180.

The ratio of budget to numbers of members in a consortium also varies a great deal.

Eighty-eight consortia furnished the data needed to calculate this ratio in 1991 and 68 revealed the data in 1996. In 1991 the range was from $472 to $921,000, with a mean of $101,000, a median of $21,000, and a mode of $25,000. When the 4 consortia with the largest ratios were removed from consideration, the mean fell to approximately $81,000. In other words, in a typical consortium whose budget is divided by the number of its members, the mean is more than $80,000.

In 1996 the range was from $103 to more than $4 million with a mean of $157,000, a median of $20,974, and a
mode of $30,000. When the 3 consortia with ratios over a million dollars each were excluded from the calculations, the mean fell to $62,900.

**Directors**
The titles of the consortia directors have changed very little over the years; however, the title “president” is more popular now than earlier.

Twenty-three different titles were listed in the Directories. “Executive Director” was the most often listed title—41 percent in 1983, 40 percent in 1991, and 46 percent in 1996. “President” was used by 19 percent of directors in 1983, 24 percent in 1991, and 26 percent in 1996. “Director” was the next most popular designation for 14 percent in the earliest survey, 13 percent in 1991, and 8 percent in 1996. Twenty other titles were listed once or twice in the surveys.

The gender of directors shows considerable change between the 1980s and 1990s.

In 1983, 81 percent of directors were men and 19 percent were women. In 1991, the percentages were 68 percent men and 32 percent women. By 1996, 67 percent were men and 33 percent were women.

The longevity of the careers of consortium directors appears to be short lived.

Only 12 names (29 percent) were listed in all three Directories—7 men and 5 women. Between 1983 and 1991, 34 names (24 men and 10 women) could be matched; 26 were employed by the same Consortium, four had been promoted in the same organization, and 4 had moved. Between 1991 and 1996, 29 names (19 men and 10 women) matched. All but one worked at the same Consortium—one moved because of his Consortium’s demise.

**Activities**
Consortia participate in a wide range of activities.

Listed in the 1991 and 1996 Directories were activities in which each consortium participated. In 1991, 60 different activities were identified with 645 total participants. The range of participants was from 1 to 41 with an average of 11 consortia involved in each activity. In 1991 the activity in which consortia most often participated was cross-registration/student exchange, which was listed by 34 percent. Following in descending order were: library cooperation (31 percent), professional development (28 percent), seminars, workshops, conferences (27 percent), publications (21 percent), business/industry relationships (20 percent), research (16 percent), graduate education (16 percent), government liaison (16 percent), joint purchasing (15 percent), telecommunications (14 percent), public relations (14 percent), minority access (12 percent), joint academic courses/programs (12 percent), international programs (12 percent), continuing education (12 percent), school-college relationships (11 percent), joint faculty/faculty exchange (10 percent), information service (10 percent), financial aid development (10 percent), computing—including planning and research (10 percent), and adult learners (10 percent). Thirty-eight additional activities were listed by fewer than 9 percent of consortia.

In 1996, 29 activities were reported with 834 total participants. The range was from 9 to 60 with a mean of 29 consortia participating in each activity. The most popular was seminars/workshops/conferences with 75 percent of consortia involved in this activity. Following in descending order were: professional development (63 percent), public relations (55 percent), publications (55 percent), information service (50 percent), library cooperation (46 percent), grants (45 percent), telecommunications (45 percent), joint academic courses/programs (44 percent), business/industry relationships (43 percent), cross-registration/student exchange (41 percent), research (35 percent), school-college relationships (34 percent), joint purchasing (34 percent), international programs (34 percent), continuing education (33 percent), government liaison (33 percent), adult learners (31 percent), joint faculty/faculty exchange (30 percent), policy analysis (30 percent), K-12 partnerships (28 percent), computing—including planning and research (24 percent), teacher education (24 percent), economic development (24 percent), minority access (24 percent), graduate education (23 percent), consulting network (19 percent), financial aid development/information (14 percent), and placement (11 percent).

Between 1991 and 1996, there appear to be large discrepancies in the percentages of consortia participating in various activities. Perhaps respondents were careful to give more complete information in 1996 and/or the directory editors categorized the activities differently. Whatever the reasons, it can be concluded that consortium participate in a variety of activities intended to benefit their member colleges and universities.

**Research is important to some consortia.**

In 1991, 19 consortia reported research activities. The largest consortium (n=1,500 members) purported to conduct “timely research on current issues and national trends affecting higher education.” Other descriptors given by respondents were: “original research to assist trustees” and “questions of common interest of public institutions in [state].” Some studies dealt with specific fields—medicine, biomedicine, biotechnology, water quality, energy research, and marine affairs. Also listed were market research, regional research, research foundations, research park, institutional research, inter-disciplinary research, minority fellowships, and library conversion of records.
In 1996, 28 directors reported that their consortium participated in research. Unfortunately 13 gave no specifics, but 13 furnished a few details. The largest consortium stated that it "serves as a key resource for educational policy-makers across the country." Another said "[we] gather and disseminate information about educational, professional, and operational issues, including statistics important to the profession." Other statements were: "[we are concerned] about the development of public policy at both state and national levels"; "[we] provide research and information for colleges"; "statistics are gathered, analyzed, and published on a wide array of topics, which are published in a variety of reports and policy papers." A survey service, publications, and newsletters were provided by another. Several emphasized public policy, while others assisted with economic impact studies, demographic studies, and timely assessments of critical issues.

Comparisons with Previous Research
Some longitudinal comparisons with previous studies are possible.

The following conclusions can be substantiated by comparing the results of this study with previous research.

In 1972, Lewis Patterson described the inadequacy of data on interinstitutional cooperation. Twenty-seven years later there continues to be a paucity of research on consortia except for studies that have been greatly delimited.

Neal (1984) stated that consortia varied greatly in composition, mission, organization, and activities, and described three approaches, including ad hoc or special-purpose consortia, national interest groups, and general-purpose consortia, both regional and metropolitan. The conclusions of this study would support Neal’s statements, because the consortia in the 1990s can still be characterized in much the same way.

In her paper delivered at the Association for the Study of Higher Education conference, Holland (1990) commented that diverse funding for consortia was necessary over the long term. The findings in this study would support Holland’s statement, because most consortia relied upon two or more sources of funding.

Discussion
Based on the results and conclusions of this study, several comments and questions can be generated. First, it would appear that the total number of consortia is not well known. The Directories published by the Association for Consortium Leadership (ACL) are incomplete. The 1996 Higher Education Directory lists 116 consortia, but only 44 could be matched with those in the 1996 ACL Directory. In other words, there were 72 additional consortia listed in the Higher Education Directory that did not appear in the 1996 ACL document and there were 36 additional consortia described in the 1996 ACL Directory that did not appear in the Higher Education Directory. Of course it is quite possible that the editors of the most recent ACL Directory did not receive responses from the 72 consortia that were not listed in their publication. The question becomes, who has a complete listing of all existing educational consortia?

Second, it is likely that the titles of many educational consortia may be confusing. Without consortium descriptors in their names, how can these organizations be easily recognized? For example, the Southern Illinois Collegiate Common Market (SICCM) is well-known in its region, but its title would not identify it as a consortium to outsiders. Likewise, the University City Science Center in Philadelphia would probably not be recognized as a consortium outside Pennsylvania.

Third, questions can be raised about appropriate education and experience for consortium directors. Why is there so little longevity in the profession for these people? Perhaps research should be conducted with directors to learn their educational backgrounds and career ladders.

Fourth, the demise of some interinstitutional arrangements is occurring. Why are consortia closing? The reasons may include the lack of a common mission and a reluctance to support consortia financially.

Finally, what IS the future of multipurpose consortia in higher education?

References


Poland, Mark W. 1986. Factors Associated with Statutory Consortium


AACRAO's SEM IX:
Strategic Enrollment Management Conference

Optimize Recruitment, Enrollment, Retention, and Graduation Rate

The SEM Conference is the premier educational experience for enrollment managers as well as other professionals seeking an enhanced understanding of the SEM discipline. The AACRAO SEM Conference is unique in that it provides participants with expertise that can be immediately translated into an action plan specific to their campus environment. Decision-makers from institutions just beginning to develop a SEM plan or improving an existing one attend the Conference. Participants engage in pre-conference readings, active learning during the conference, case study assignments, and networking opportunities. The goal of the SEM Conference is to customize the learning experience for each individual and in so doing, provide value that exceeds the financial and time investment of attending.

November 10-14, 1999
(dates include optional pre- and post-conference workshops)

Disney's Coronado Springs Resort
Lake Buena Vista, Florida (outside Orlando)

- best practices in SEM
- SEM templates and resources
- case studies
- institution-specific learning plans
- pre- and post-conference workshops
- mentoring from SEM experts
- networking with SEM professionals

SEM IX: A different kind of learning experience...

For more information:
www.aacrao.com • meetings@aacrao.nche.edu • (202) 293-9161
Level of Math Preparation in High School and Its Impact on Remedial Placement at an Urban State College

Jeff E. Hoyt

Background

Colleges and universities have begun to focus more attention on high school programs in their geographical area because of the high remediation rates of entering freshmen. There is an interest in the reasons for the low math placement test scores of entering freshmen and how they relate to a student’s math preparation in high school. The result of such research has created more interest and support for joint efforts between institutions of higher education and local high schools to improve education or articulation.

Montgomery College in Maryland developed a partnership with local high schools to track student transfer from high school through college creating a data file with each student’s high school and college transcript data. The College reported that 60 percent of the students who completed the placement test needed remedial education. They found that “nearly 54 percent of the . . . students entering the College for the first time . . . had completed Basic Math . . . as their highest level of math” (Office of Institutional Development 1994). Another 34 percent of the students reported that Algebra I or Beginning Algebra was their highest level of math completed in high school (Office of Institutional Development 1994).

These low rates of math preparation for students graduating from local high schools require institutions of higher education to invest substantial resources into remedial education programs. Colleges and universities have initiated joint efforts with local high schools to improve the education of students by revising college education programs for teachers, developing technology to enhance the curriculum in local school districts, sharing information and analysis, and establishing committees to address educational issues (Justiz 1997; Garies and Larsen 1997).

Focus of this Study

Although prior research at the secondary level provides important insights into the math preparation of students, the present study takes research at the secondary level a step farther by relating the level of math preparation in high school to remedial placement and subsequent performance in math courses offered at

Jeff E. Hoyt is an adjunct faculty member and research analyst at Utah Valley State College. He was a research analyst for the Utah Commissioner’s Office, and is a doctoral candidate at the University of Utah.
the college level. This study was initiated as a joint effort by Utah Valley State College (UVSC) and a local high school district to find ways to improve the math preparation of high school students.

The study found that recent high school graduates who successfully completed more advanced math courses in high school had higher ACT math test scores, higher math placement test scores, and were less likely to take remedial courses at the College. UVSC enrolled students with various levels of math ability ranging from graduates who completed only remedial math courses in high school, to those who successfully completed Calculus. About 35 percent of the high school students who graduated from the district during 1995-1997 did not successfully complete Algebra 2 or Intermediate Algebra. This course was needed to avoid placement in remedial coursework. However, there was still a substantial remedial placement rate for students completing Intermediate Algebra and Honors Intermediate Algebra. This raised concerns about the rigor of high school math programs, grading policies, and the ability of students to retain math skills learned in high school. While some students may not give the testing their best effort, nevertheless, the percentage of students testing into remedial courses substantially declined when they successfully completed College Algebra, Pre-Calculus, Honors Pre-Calculus, and Calculus.

The study also found differences in the remedial placement rates for each high school. Differences in the remedial placement rates among high schools were explained primarily by eighth grade SAT math scores and to a lesser extent by gender, rather than by attendance at a particular high school. Students in math courses at some high schools may be better prepared for the course than are students taking the same course at other high schools. Instructors with better prepared students may be able to cover the course material in more depth, reducing the remedial placement rate for students who later attend college.

The following sections begin with an analysis of student high school math preparation and performance on the ACT test, followed by a profile of graduates from the district attending the College. The level of math a student successfully completed in high school and its impact on remedial placement at the college is then examined, with a discussion of the need for actual mastery of math skills rather than allowing students who have not learned the material to receive passing grades. This is followed by further analysis of differences in remedial placement rates for graduates from different high schools. The implications of these findings are then discussed.

Methodology

The data included information from the high school and college transcripts of recent high school graduates of a local school district who later attended UVSC. These students graduated between 1995 and 1997 and attended the College by fall 1997. The students' highest level of math completed in high school and first math course completed at the College were identified, along with the students' grades in the courses. The grades for each quarter in high school were then averaged for analysis.

Several variations in the titles of math courses were entered on student transcripts. A master table of each title and a standardized title were created for analyzing the data. The math curriculum specialist in the district was consulted to develop the standardized titles. A small number of the math titles such as "Algebra" could not be used to identify the level of math completed by the students. If this was the highest level of math completed by the students, they were excluded from the analysis. Students with college transfer credit were also excluded from the study.

Students were classified as successfully completing a high school course if they completed one credit hour or year of instruction and received a "C-" or higher in the course. High school course descriptions and the district office confirmed that all math courses at the high schools were taught over a full school year for one credit.

Logistic regression was used to identify the factors that predicted placement in remedial education at the College. Four high schools in the district were used in the analysis. The regression compared three high schools with substantially higher remedial placement rates at the College with a fourth high school with lower remedial placement rates. The analysis controlled for gender, ethnicity, eighth grade math test scores, and cumulative high school GPA. Statistical significance was set at a p < .05.

High School Math Preparation and ACT Test Performance

The performance of high school graduates on the ACT test was analyzed to determine if a student's level of math achievement in high school was related to math scores on the ACT. Many students took the ACT before completing their last math course in high school because they needed to meet application deadlines for college. Only some of these students took the test again after completing their last math course. Because of this, the performance of many students on the ACT test may have been higher had they taken the test after fully completing their last math course during their senior year.

The results indicated that as students successfully completed higher levels of
Math, they earned higher ACT math scores on average. Students completing Introductory Algebra and Geometry courses in high school had an average ACT math score of 17; whereas students who completed Calculus had an average ACT math score of about 23 (see Table 1).

Math Preparation of Students Attending UVSC

Because the College has an open enrollment policy, the general level of math achievement in local high schools directly impacts the resources that are spent on remedial education at the College. The College operates a testing and remedial education program to identify students who need remedial courses; these courses give students a second chance at mastering the math skills requisite to succeed in college. For the last five years about half of all entering freshmen required remedial education at the College.

The transcript data for students from the district indicated that students with a wide range of high school math preparation attended the College. For the 1995 cohort that enrolled at the College, 69 percent successfully completed Intermediate Algebra or higher, 24 percent successfully completed Pre-Calculus, and 11 percent successfully completed Calculus. For the 1996 cohort that enrolled at the College, 68 percent successfully completed Intermediate Algebra or higher, 26 percent successfully completed Pre-Calculus, and 10 percent successfully completed Calculus. Complete data were not available on the 1997 cohort for the study.

Intermediate Algebra is needed to enroll in College Algebra at the College. If students lack this preparation, they will almost always require remedial education. About 31 to 32 percent of the graduates who subsequently enrolled at the College from 1995 through 1996 did not successfully complete Intermediate Algebra in high school, which increased the need for remedial education on campus.

High School Math Preparation and Placement in Remedial Education Courses

The Learning Enrichment Center (LRC) basically offers two remedial math courses for students: Foundations For Algebra, and Introductory Algebra, although a small handful of students will take Individualized Math Fundamentals. Placement in Intermediate Algebra is not considered to be remedial education at the College. These courses cover signed numbers, polynomial expressions, factoring, quadratic functions, and other basic algebra concepts.

The COMPASS math test is an adaptive computerized test that is used to place students in math courses at the College. It consists of questions in Pre-Algebra, Algebra, College Algebra, and Trigonometry. The computer program moves students back and forth among these tests in a seamless transition, depending on their performance on each test. Only one of the tests results in a valid score for course placement.

Because math scores from five different tests, including ACT math, Pre-Algebra, Algebra, College Algebra, and Trigonometry, were used to place students, the analysis of high school math preparation and test performance could not be done with actual test scores. The highest placement recommendation, regardless of the test, was used in the analysis. Course placement recommendations are based on the criterion of earning a “B” in math courses at the College.

According to guidelines at the College, students who attended UVSC and had a math score of 24 on the ACT could take College Algebra, and students with an ACT math score of 19 could take Intermediate Algebra. Students with ACT math scores ranging from 16 to 18 were placed in Introductory Algebra, and those with a score below 16 were placed in Foundations For Algebra. Students who wanted to take Calculus needed a COMPASS test score.

Although the level of math completed in high school improved a student’s test performance, a large percentage of the graduates who completed advanced math courses in high school still tested at the remedial level (see Table 2). This raised concerns about the rigor of high school

<table>
<thead>
<tr>
<th>Table 1: Level of High School Math Completed and ACT Math Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Level Of Math Completed</td>
</tr>
<tr>
<td>Introductory Algebra, Geometry</td>
</tr>
<tr>
<td>Intermediate Algebra</td>
</tr>
<tr>
<td>Pre-Calculus, College Algebra</td>
</tr>
<tr>
<td>Calculus</td>
</tr>
</tbody>
</table>
math programs, grading policies, and the ability of students to retain math skills learned in high school. Also, some students may not give the test their best effort.

A large percentage of students completing advanced math courses in high school tested into remedial courses at UVSC: Foundations for Algebra or Introductory Algebra. About 62 percent of graduates completing Intermediate Algebra, 46 percent of graduates completing Honors Intermediate Algebra, 24 percent of graduates completing College Algebra, and 29 percent of graduates completing Pre-Calculus tested into remedial courses when they attended UVSC.

### High School Math Preparation and Success in College Math Courses

Students at the College often needed to repeat the same level of math or take a lower level of math than the final course they took in high school to maintain a "B" average in college math courses. The average grade in most math courses at the College was about a "C+." Graduates who did not follow the placement recommendations generally performed less well, and often failed courses when they attempted to bypass more than one remedial math course. Graduates who took Intermediate Algebra in high school, tested into Foundations for Algebra, and then took Intermediate Algebra at the College on average earned a 1.97 GPA in the college course. Graduates who took Intermediate Algebra in high school, tested into Intermediate Algebra at the College on average earned less than a "C" in the college course. If students bypassed one remedial course, and took Intermediate Algebra instead, their grade in Intermediate Algebra at the College was on average about a "C+" whereas students prepared for the course on average earned a "B" in the course. In other words, students should follow the placement recommendations at the College that are based on their ACT or COMPASS math test scores if they want to avoid failing courses or if they want to maintain a "B" grade on average in math courses. This often requires students to repeat math courses or take a lower level of math in college than completed in high school.

### Differences Among High Schools in Student Performance

A final analysis was conducted to assess whether some high schools in the school district were doing a better job of preparing students in math. It was felt that the district could use this information to find out what the most successful high schools were doing and possibly incorporate changes in high schools that were not preparing students as well in math. Graduates who successfully completed Algebra 2 or Intermediate Algebra in the high school and subsequently enrolled at the College were the only group of students that was large enough to analyze by high school.
The initial analysis indicated that there were substantial differences by high school in the math achievement of graduates in this group (see Table 3). This comparison among high schools in the district was made on the following assumptions. If high school students take Intermediate Algebra in high school, and earn a grade that represents a certain level of competency, then there should not be great variation among schools in the math performance of these students and their subsequent remedial placement. As can be seen in the tables, High School B and High School E had the lowest percentage of graduates who tested into Foundations For Algebra at the College. High School A and High School D had about twice as many graduates testing into Foundations For Algebra at the College than High School E. High School B and High School E had more students testing into Intermediate Algebra. High School E had substantially more students testing into College Algebra.

Logistic regression was used to assess whether there was a significant difference among the high schools in their level of remedial placement (see Table 4). A dummy variable was used (1 = Remedial Placement and 0 = College Level Placement) to identify students who tested into either of the remedial math courses at the College. Dummy variables were also used for gender (1 = Female, 0 = Male), and ethnicity (1 = Hispanic, American Indian, Pacific Islander, 0 = White, Asian). Eighth grade math scores were measured using the students' percentile rank on the SAT test completed in the eighth grade. The cumulative high school GPA was also included.

### Table 3: Intermediate Algebra (C- or higher) and Recommended Placement By School

<table>
<thead>
<tr>
<th>School</th>
<th>N</th>
<th>Foundations for Algebra</th>
<th>Introductory Algebra</th>
<th>Intermediate Algebra</th>
<th>College Algebra</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School A</td>
<td>60</td>
<td>33%</td>
<td>35%</td>
<td>24%</td>
<td>9%</td>
</tr>
<tr>
<td>High School B</td>
<td>30</td>
<td>15%</td>
<td>37%</td>
<td>44%</td>
<td>4%</td>
</tr>
<tr>
<td>High School C</td>
<td>160</td>
<td>23%</td>
<td>41%</td>
<td>26%</td>
<td>9%</td>
</tr>
<tr>
<td>High School D</td>
<td>100</td>
<td>28%</td>
<td>44%</td>
<td>20%</td>
<td>8%</td>
</tr>
<tr>
<td>High School E</td>
<td>86</td>
<td>13%</td>
<td>32%</td>
<td>38%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Note: Percentages may not add up to 100% because of rounding.

### Table 4: Logistic Regression—Remedial Placement for Students Completing Algebra 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1.0330</td>
<td>.2952</td>
<td>12.2469</td>
<td>1</td>
<td>.0005</td>
<td>.1556</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>.1761</td>
<td>1.0544</td>
<td>.0279</td>
<td>1</td>
<td>.8674</td>
<td>.0000</td>
</tr>
<tr>
<td>Eighth Grade Math Scores</td>
<td>-.0554</td>
<td>.0077</td>
<td>51.7296</td>
<td>1</td>
<td>.0000</td>
<td>-.3428</td>
</tr>
<tr>
<td>Cumulative HS GPA</td>
<td>-.4098</td>
<td>.4088</td>
<td>1.0050</td>
<td>1</td>
<td>.3161</td>
<td>.0000</td>
</tr>
<tr>
<td>High School A</td>
<td>.4381</td>
<td>.4563</td>
<td>.9218</td>
<td>1</td>
<td>.3370</td>
<td>.0000</td>
</tr>
<tr>
<td>High School C</td>
<td>.3847</td>
<td>.3418</td>
<td>1.2665</td>
<td>1</td>
<td>.2604</td>
<td>.0000</td>
</tr>
<tr>
<td>High School D</td>
<td>.7533</td>
<td>.4069</td>
<td>3.4275</td>
<td>1</td>
<td>.0641</td>
<td>.0581</td>
</tr>
<tr>
<td>Constant</td>
<td>4.4522</td>
<td>1.4722</td>
<td>9.1451</td>
<td>1</td>
<td>.0025</td>
<td></td>
</tr>
</tbody>
</table>

Spring 1999
High Schools A, C, D, and E were included in the analysis. High School B was excluded because of small sample sizes. Attendance at High School E was used as the comparison group because of its lower remedial placement rate than the other high schools. Dummy variables were used for High Schools A, C, and D (1 = attended the high school, 0 = did not attend the high school). The comparison among the high schools was whether attendance at High Schools A, C, or D increased remedial placement when compared with High School E.

When there were no control variables, attendance at High Schools A, C, and D had a significant and positive impact on increasing remedial placement at the College (probability of .05 or less) when compared with attendance at High School E. When controlling for the other variables, statistical analysis indicated that the differences in the remedial placement of graduates was explained primarily by eighth grade math test scores, and to a lesser extent by gender. When the control variables were used, differences in remedial placement attributed to attendance at a particular high school became insignificant.

It appears that High School E, with the lowest remedial placement rate, had students with greater aptitude in math or better prepared students as reflected by their higher eighth grade math test scores as a group. These students may have been able to cover algebra concepts in more depth in the high school course, resulting in lower remedial placement rates at the College.

Overall, 78 percent of students who tested into remedial education at the College were correctly classified by the model. About 86 percent of students testing into remedial courses were correctly classified, and 63 percent of students who did not test into remedial courses were correctly classified. The chi-square test was significant which resulted in a rejection of the null hypothesis that the coefficients of the variables in the model were zero (with the exception of the constant).

**Study Implications**

The study results indicate that the level of students' math preparation in high school makes a substantial difference in their performance on the ACT test, math placement tests, and in their ability to take college level math courses rather than remedial education. The present study also raises concerns about the rigor of some high school math programs. High schools may be passing students on to take advanced math courses, when in fact the students have not adequately mastered the material in prerequisite courses.

The study also raises concerns for colleges and universities that must provide remedial education for students. Remedial education provides a second chance for students; however, it is also costly for colleges and universities, state government, and ultimately the public and students who must pay for remedial math courses—whose contents they should have mastered in high school. Students may come to college believing that they have adequate math skills to take college level math when in fact they have not sufficiently learned prerequisite material. They may decide to ignore math placement recommendations and take a higher level of math in college, thereby risking lower grades and possibly failing grades. Another explanation may be that students are not retaining the material by the time they take math courses in college. College courses may be more rigorous and require a better mastery of concepts in prior courses. It is also possible that some students may not give the testing their best effort.

The local high schools can help improve the math preparation of students for success in college in several ways. The local high schools currently require students to complete four years of English. To improve the math preparation of students, districts could require four years of mathematics instruction. Students should be advised to complete at least Intermediate Algebra to substantially reduce the need for remedial math instruction in college. However, students should be strongly encouraged to complete College Algebra or Pre-Calculus. High schools need to ensure accountability in that grading policies represent comprehension and mastery of math skills rather than social passing along of students who have not sufficiently learned the basics.

The local high schools have an applied math sequence, taken over two years, that is the equivalent of Introductory Algebra. A slower paced program could also be developed or used more widely for Intermediate Algebra. If students take a class, and the instructor becomes aware that the students are really not prepared for the course, the instructor could redirect these students into the slower paced course.

It may be more effective to redirect students in math courses in junior high school before they enter high school. In other words, instructors in junior high school could redirect students who are not sufficiently prepared into math courses more appropriate to their skill level. In this way, a passing grade in a course may represent competence in the math skills, rather than moving along students who are making an effort but who are not fully learning the concepts.

A student's ability to obtain a degree in college is heavily impacted by the ability to learn college level math skills. Successful completion of College Algebra is required for bachelor's degrees, and Intermediate Algebra is required for associate degrees. An inability to complete college level math may hinder students from successfully earning their degree.
Students may become discouraged when they have to take several remedial math courses. A lack of math skills also limits a student's opportunities or choice of major.

Colleges and high schools benefit from developing cooperative agreements to share information for research purposes to inform local high schools about their effectiveness in teaching students math skills and preparing them for college. It is our hope that these efforts will improve the math preparation of high school graduates and their subsequent success in college.

References


The Forum

Commentary

Why We Can’t Break Out of the Box

David James

Abstract

The recent literature in higher education encourages a radical new approach to the way colleges and universities do business. Change is the platitude proclaimed as the new millennium arrives. However, the existing culture and structure of organizations prohibit the magnitude of change required. Unless sweeping, substantial change guidelines are in place, people will subvert the change process and continue to seek routine, habit, and security in their work environments.

The current hot issues in higher education—technology, accountability, decreasing enrollments and funding, political intervention and control, and outcomes-based assessment—all signal the need for change. This need has been eloquently summarized by Patricia Carter and Richard L. Alfred in a publication titled Breaking Out of the Box: New Colleges for a New Century (1996). The authors summarize the future themes for colleges with the following statements:

(Colleges could) become borderless institutions focused on speed by turning the entire organization into autonomous business units that operate at a considerable distance from executive management.

(Colleges could) organize educational delivery systems around just-in-time principles where courses and services could turn on a dime to accommodate students’ needs and preferences while maintaining their academic credibility.

(Colleges could) turn large parts of the organization into a “virtual experience” no longer dependent on permanent facilities by partnering with outside organizations to put courses, programs, and services on the Internet.

(Colleges could) increase productivity of academic departments by assigning strategic management responsibilities (planning, assessment, marketing) to faculty as a routine part of workload or outsourcing instruction altogether.

(Colleges could) adopt financial practices from the corporate sector which rely on new and recently developed programs to produce a third or more of operating revenue in any given year and assign a significant portion of the budget as “risk capital” (Carter and Alfred 1996).

While I wholeheartedly agree with the ideas presented, and admire Carter and Alfred for challenging the higher education community with such innovative material, I frankly don’t see such leaps of change forthcoming. As a dean in a community college system with 25,000 students, I live in a world that views change largely with suspicion, derision, and, sometimes, with outright anger. How can we expect faculty and staff to “break out of the box” when it’s difficult to get them to move “inside the box”? Mundane educational concepts like weekend colleges, credit for life experience, and accelerated class offerings are met with strong resistance. Many union staff recoil at the thought of taking on new and evolving duties and instead hold the administration to the letter of the current contract. Change is seen more as a subversive agent rather than as the opportunity for growth and renewal.

The Desire for Security

If one ascribes some truth to Abraham Maslow’s Hierarchy of Needs theory, then the human response to change like this can be partially explained. As Maslow states, before people can move upward to higher levels of achievement, growth, and productivity, the lower level needs of shelter, food, and safety must be satisfied. The need for safety, which includes order and stability, is an overriding concern for people who are in the midst of structural and individual change. Will I still have a job? Will I be asked to take on new roles, and will I enjoy them? Will I receive adequate training? What happens to me if I fail? The questions outweigh the answers in times of transition.

David James works as a Dean of Academic and Student Services for Oakland Community College in Michigan. He received his doctorate in curriculum and instruction from Wayne State University. His most recent chapbook of poetry, Do Not Give Dogs What Is Holy, was published by March Street Press in 1994.

Vol. 74, No.3 44
An Issue of Commitment

For any organization to succeed, employees must be dedicated and committed and willing to work effectively beyond the scope of their individual jobs. However, asking employees to embrace change and commit themselves to charting a path into unknown territory is not only one-sided, but foolish. Most people will seek out routine, habit, and security over radical change. Most people will subvert the process of change within an institution if they perceive change as a threat to their livelihood.

To effectively engage workers in change, the organization must 1) commit itself to the individual, 2) create incentives to change, and 3) establish an environment that expects and accepts a reasonable level of failure.

A Commitment to the Individual

Employees will be more likely to leap into the dark if they are reassured that they will keep their jobs. A promise to the employees' future will lessen the fear of change and set the tone for redefining work. Most workers are competent and want to grow and learn. The organization must guarantee the training required to be successful employees in this new world. This commitment to the individual must be as strong as the organization's commitment to change. If not, the individual will draw back and even work to undermine the process.

Incentives to Change

The organization must also provide legitimate reasons for people to alter their behaviors and attitudes. Beyond the theoretical commitment to change as a growth process, a reward system based upon innovation and creativity is essential. The culture within the organization has to expect and encourage change attempts. As most businesses will attest, if one's goals are far more likely to be reached.

An Environment for Change

Essentially, higher education employees are presented with this scenario:

Worker: OK, I'm willing to change. Tell me what you want me to do.

Organization: We don't know. We want you to do things we've never done before, and we're not sure what those things are.

Those in academe pride themselves on being right. Few professors or staff enjoy their own failures; in fact, they have created a whole life around the concept of being right. Employees must be given the overt permission to succeed, and fail. As the old Japanese proverb states, "Fall seven times, get up eight." The institution must provide an assurance like a safety net as it asks people to venture out on the limb of change. The environment must allow for mistakes and disasters as people journey into unfamiliar waters, first learning how to float, and then to swim.

Out of the Silos

The consistent, alluring refrain echoed by many reformers of the world today can be stated aptly: "The shape of colleges in the future depends, in large part, on the ability of leaders to look at the institution in a new way, through a new lens" (Carter and Alfred 1996). Colleges must "break out of the box" and embrace innovation and creativity in a new learning environment. Colleges must encourage faculty to step out of their "silos," out of their departmental and territorial traditions, and engage in dramatically new roles.

However, the reality of the situation indicates that if people jump completely "out of the box," they will leave most of the others behind. In reality, institutions cannot exert enough force at one time to move "out of the box." And, in fact, the autocratic process currently used to coerce people to change only fuels resistance.

For our future discussions on educational reform, a more appropriate metaphor may be to urge colleges and faculty and staff to get one foot out of the box and limp along, dragging the box forward until the edges break. If an organization can commit itself to its workers and create incentives and an environment for reasonable change, then the seed for innovation has been planted. The box will eventually fall apart, leaving behind pieces to be written about by historians. This process may take months; it may take years.

As history books show, unless change encompasses the majority, it will not last long.

Reference

For admission to our graduate programs, we require "a bachelor's degree or equivalent from an accredited college or university." Our policy does not specifically say "regionally accredited," but that is the usual interpretation.

The Ministry of National Education and Religion in Greece does not recognize any private institutions in Greece as universities. Thus they all seem to be the equivalent of "non-accredited." However, one private college in Greece is accredited by the New England Association of Schools and Colleges in the United States. Another is not accredited by any regional accrediting association in the United States, but it is accredited by the Accrediting Commission on Independent Colleges and Schools (ACICS), which is recognized by the U.S. Department of Education as an official accrediting agency.

How should we deal with these types of institutions? Rely upon the status given to them by educational authorities in their country? Or pay attention to their U.S. accreditation?

A.

Accreditation by one of six regional accrediting associations is the process whereby an educational institution in the United States obtains the official recognition as a degree-granting institution that merits reciprocal acceptance of its courses and degrees by other educational institutions. The other types of accrediting organizations in the United States confer a similar status, but one that is not universally accepted in the United States.

Each university and college in the United States is autonomous. Each can set its own policies concerning the types of educational institutions from which it will accept courses and degrees. Each can waive its own policies whenever that action is deemed appropriate.

When most universities and colleges in the United States say "accredited," they are referring to regional accreditation. However, admission and transfer of credit policies are not always stated clearly, and they are not always applied uniformly. Some institutions accept courses and degrees only from a regionally accredited institution. Some waive that policy in limited cases. Some accept courses and degrees from institutions with other types of accreditation. Some accept courses and degrees from institutions that have no accreditation, especially if the other institutions are sponsored by the same religious tradition. The situation becomes more complex when the teaching institution is located outside of the United States.

Most educators in the United States consider the foreign equivalent of regional accreditation in the United States to be official degree-granting recognition granted by the officials who have jurisdiction over tertiary (university-level) education in the country in which an institution operates.

Some institutions are not officially recognized as degree-granting institutions in the country in which they operate but have been accredited by one of the regional accrediting associations in the United States, or by one of the other types of accrediting organizations in the United States. If a university in the United States accepts courses and degrees from another educational institution in the United States only if it has regional accreditation, it would logically not accept courses and degrees from these types of institutions, because they do not have the equivalent of regional accreditation from the educational authorities of the country in which they operate.

Most universities and colleges in the United States however, do accept courses and degrees from an educational institution in another country that has been accredited by one of the six regional accrediting associations in the United States, even if the institution has not received official degree-granting recognition in the country in which it operates. While this might make some sense within the context of the U.S. educational system, accreditation by a U.S.-based organization does not affect the policies of universities in other countries, which almost uniformly do not accept courses and degrees from this type of institution.

This dichotomy can cause problems for a student who returns home after receiving a bachelor's degree from a U.S. institution that
accepted a large number of credits in transfer from an institution that is not officially recognized as degree-granting in that country. It can cause even greater problems when a student has received a master’s degree or doctoral degree from a U.S. educational institution but does not have a bachelor’s degree from an institution that is officially recognized as degree-granting by educational authorities in the student’s home country. If an applicant’s bachelor’s degree is not officially recognized, the master’s degree or doctoral degree is tainted.

Q.

An applicant completed a two-year Rehabilitation Technician course at a training school in Zimbabwe. The Health Professions Council of Zimbabwe issued the final examination certificate.

The student’s current U.S. university did not grant any transfer of credit for this program. One of my sources says that credit may be given for “technical college diplomas.” Is a technical training school outside of the technical college definition? Is that why the student was denied credit?

Do we ever void each other’s transfer of credit decisions and give credit for courses or programs for which another U.S. institution has denied credit?

A.

Because each university and college in the United States is autonomous, each one sets its own requirements for admission and its own requirements for graduation. Each one determines the compulsory and elective components of its own degree programs.

At many institutions, each division is autonomous. For example, a course that can be used to meet some of the requirements for a degree in a college of letters and science might not be acceptable for meeting any of the requirements for a degree in a college of engineering, and vice versa.

These differences in policies, and additional ones, apply to courses taken at another institution. Again, each institution sets its own standards for transfer of credit, and within some institutions, each division sets its own standards.

It is appropriate for these differences in standards to exist within the educational system of the United States. Standards are set in the context of each educational program’s purposes and goals. One of the strengths of our educational system is the flexibility of standards across the system. We do not have a monolithic homogeneous university sector. There are no national standards. No public or private agency in the United States has the authority to set national standards. Because each educational institution in the United States sets its own standards, it would be inappropriate for it to simply copy whatever policy another educational institution might have. Each policy should be developed within the institution, based upon its own educational purposes and goals.

It is quite likely that the policies which one U.S. university applies to courses offered by an educational institution in the United States will not be the same as the policies that another U.S. university applies.

The same kinds of differences are appropriate when an educational institution in another country is involved.

The task of a foreign student admissions officer is to identify the nature of an educational institution in another country, and the nature of the specific educational program in which a foreign-educated applicant has been enrolled; then to identify the U.S. equivalent of both the institution and the program; and then to apply to courses taken in that program the same policies that would be applied if the program had been taken at an educational institution in the United States.

One university in the United States might decide not to grant any transfer of credit for a program completed at a training school in Zimbabwe in conjunction with the Health Professions Council of Zimbabwe because neither the training school nor the Council is an officially recognized degree-granting institution in Zimbabwe, and the U.S. institution only grants credit for courses completed in the United States if the courses were taken at a regionally accredited institution.
Another university in the United States might decide to grant transfer of credit for the courses completed in Zimbabwe because in its judgment the training school and the program are the equivalent of a health training program offered by an educational institution in the United States for which it does grant transfer of credit.

Technically, you were incorrect when you stated that "one of my sources says that credit may be given for technical college diplomas."

There is no published source that tells any educational institution in the United States what it may or may not do in making a decision concerning admission or transfer of credit. There are publications that include recommendations. These recommendations represent the judgment of the people who wrote and approved them at the time that the publication was being prepared.

Recommendations are just recommendations. They give a foreign student admissions officer a base point from which to work in determining which of an institution's own policies it is appropriate to apply to the foreign educational qualification(s) involved.

It is the pedagogical responsibility of each university and college in the United States to be true to its own educational purposes and goals. Because we have a heterogeneous system of autonomous institutions, not a homogeneous system, policies will differ, and students who seek to transfer from an institution may meet differing decisions at others. This is quite appropriate. It is the student's responsibility to transfer to the educational institution whose policies and degree programs best meet the student's educational purposes and goals.

Q.

How should we view courses completed solely through distance learning?

A.

In the 1970s and earlier, many universities in the United States did not grant transfer of credit for courses completed through correspondence (an early form of distance learning), even if the correspondence courses were offered by a regionally accredited U.S. university. Now few institutions discriminate against specific forms of delivery of higher education. If (and that is a very big word) the standards of performance are the same, it should not matter how or where a student learned.

When he was chairman of the Williams College board of trustees, James A. Garfield (later a U.S. president) said that his idea of a perfect educational setting was Mark Hopkins on one end of a log and a student on the other. (Mark Hopkins was the president of Williams College at that time.) We now have electronic logs that create larger distances between teacher and student than one finds in a classroom, but shorter distances than in courses taught via correspondence.

The question today is not "how was the student taught?" but rather "what did the student learn?" If you trust the institution that measured the quantity and quality of learning, how the student was taught is not important. If you do not trust the institution, then how the student was taught is still not important.

—James S. Frey
President
Educational Credential Evaluators, Inc.
The Shape of the River

By William G. Bowen and Derek Bok

Princeton University Press, 1998
472 pages; $24.95

Former Presidents Bowen (Princeton) and Bok (Harvard) and their associates have exhaustively explored phenomena about which there has been much conjecture and argument but little solid data. By examining data concerning the admissions, graduation, and eventual achievements of African American students at over two dozen selective colleges and universities, Bowen and Bok have provided a basis for rich discussion and some conclusions worthy of the most serious consideration. Their project features substantial data covering an extensive period of time, and rather sophisticated analyses have been included. Of special note is the "testing" of multiple hypotheses; e.g., what might have been the result of the same African American students having attended less selective institutions.

The Shape of the River is subtitled "Long Term Consequences of Considering Race in College and University Admissions." The authors clearly meant to supplant the discussion of the educational pipeline and the "enrollment management funnel" with the more accurate metaphor of a river. They even more wisely chose Twain's colorful discussions of the complexity and vagaries of the Mississippi River. What natural feature more fully bespeaks the American experience than does the Mississippi?

Admissions officers, registrars, and others familiar with phenomena that bring students to higher education are likely to embrace easily the notion that students do not flow via a pipeline with precise valves and faucets that leads qualified students to college classrooms. The multiple paths that lead to even the most selective institutions, where everyone is a "full-time" resident student, are more like a river with hundreds of tributaries, locks, eddies, cascades, and occasionally stagnant pools. We can at best over a lengthy period of time, only modify the shape of the river. Even the most prestigious university or college cannot manage with precision the plumbing that brings successful students to their campus, and large but short-term modifications may yield unintended and unfortunate long-term consequences.

On balance, Bowen and Bok argue that race conscious admissions decisions, i.e., affirmative action, have produced many of the results for which they were intended. Namely that more African Americans have enrolled at the twenty-eight institutions included in their research and that these students have graduated in acceptably high proportions. Moreover, these graduates have achieved well in terms of careers, earnings, job satisfaction, and civic participation as well as other factors. They also seem to argue rather convincingly that the success of these African Americans is potentially greater than it might have been had these college graduates attended less selective institutions. (Admissions officers of less selective institutions will justifiably take issue with such assertions.)

The above being the case, affirmative action admissions as described in the Bakke decision seems to be working, and its demise, e.g., Hopwood and Proposition 209, appears unwise. Considerable inequality of opportunity still exists, and affirmative action admissions should be continued. The authors also make a convincing case that alternative criteria such as socio-economic status or lesser dependence upon test scores are unlikely by themselves to lead to increased diversity among the enrollments of selective institutions. Therefore, Bowen and Bok seem to conclude that affirmative action admissions need to continue at selective institutions until such time as all current minorities can feel and be "unselfconsciously included."

The authors' arguments are convincingly made and elegantly supported with data. A treasure of information has been provided, and many of us already kindly disposed to race conscious admissions will fuel bulwark for our defense of what seems to now be at least partially prohibited in California, Washington, Texas, Mississippi, and Louisiana. Persons of vision in higher education should be grateful to have The Shape of the River at their disposal. The data and the arguments are extraordinarily valuable and timely.

This book, despite its many virtues, may still fail to do more than preach a well-documented sermon to an already faithful choir. The following are a few problems that seem to confront use of Bowen and Bok in a broader arena. In what may be an eastern perspective, only data and analyses pertinent to African Americans are presented. Although important and historically significant, such analyses may be of limited usefulness in California, large portions of the Southwest, and Florida. It is, of course, possible that the Latinos and Native American experience, its benefits, and achievements are or will be similar to those of the African Americans studied by Bowen and Bok. But how can the opponents of affirmative action be convinced of the certainty of that outcome?

The twenty-eight institutions from which most of the authors' data are drawn include many of the finest institutions in this nation and the world. Unfortunately,
only four are public universities. For the most part, it is public universities that have been impacted most by the anti-affirmative action movement. To date, the anti-affirmative action movement has not seriously attacked private higher education. The admissions practices of exclusive and often strongly endowed colleges and universities are governed primarily by trustees and faculty and are influenced by alumni and other supporters of these institutions. Proposition 209 in California may not have changed anything at Stanford, and Hopwood has probably brought about no modifications of Tulane’s admissions policies.

It is unfortunate that Bowen, Bok, and their associates did not address the same analysis upon the University of California or Texas A&M, two cases where the application of race conscious admissions might have provided opportunities for similar longitudinal research. Has the practice of long term affirmative action in admissions at these universities produced more African American physicians, Latina lawyers or Native American CEOs? We probably don’t know, and someone probably needs to discover the answers to such questions. Hopefully, similarly rich databases are being built for other cohorts of colleges and universities where race conscious admissions have been used to increase diversity at selective institutions. Absent these data and the appropriate analyses, one of the most important public policy questions of our time seems likely to be unanswered.

Over forty years ago, an African American woman of my acquaintance was wont to say, “Them as has gets.” She meant that if one already has material wealth the probability of increased riches is greater. Education is still the major method by which wealth is redistributed in a free economy. If attending and graduating from a more selective college or university increases a person’s ability to lead a materially and personally more satisfying and productive life, then access to such an education should be provided by processes that consider more than just rank in high school class and test scores. We need to know more about the shape of the admissions river. It seems unlikely that race conscious admissions will be further curtailed unless additional data-based arguments can be provided. *The Shape of the River*: Long Term Consequences of Considering Race in College and University Admissions is a fine start towards providing data-based arguments for the continuation of a practice whose useful life ought not yet be over.

—James Blackburn
Director of Admissions and Records
University of California, Fullerton

**The Forum**

**Taking Women Seriously: Lessons and Legacies for Educating the Majority**

By M. Elizabeth Tidball, Daryl G. Smith, Charles S. Tidball, Lisa E. Wolf-Wendel

The American Council on Education and The Oryx Press, 1999

Everyone has a stake in the education of women. This is because women of all backgrounds constitute the largest untapped source of talent in our nation, and educated women are the surest hope for the development of a humane, equitable, and productive society.

With this premise, that everyone has a stake in the education of women, Lee Tidball and her co-authors have researched and written a valuable new book that presents a plan for better educating women students at all colleges and universities. The book presents further evidence for Tidball’s earlier claim that women’s colleges produce women who achieve. But the authors’ claim in this work is not that women’s colleges are the best way of educating women. Rather, the purpose of this book is to present the successful strategies used by women’s colleges to produce high achieving alumnae so that these strategies can then be implemented at colleges and universities across the board.

Case studies are presented of a wide variety of schools with successful outcomes in educating women. The authors purposefully choose schools beyond the elite Seven Sisters, to include colleges small and large, “elite,” “selective,” and even “non-selective.” The results are surprising. The careful statistical analysis reveals that selectivity, endowment size, faculty salaries, percentage of faculty with doctorates, and quality of library holdings are not the most significant factors leading to the success of alumnae. Instead, Tidball and her colleagues’ research produces evidence that several characteristics often ignored by co-educational institutions are more influential. For example, Taking Women Seriously presents statistical evidence that the percentage of female faculty has a direct correlation to the number of alumnae that hold doctorates. Also influential, according to the data presented, is a college mission that emphasizes the importance of educating women, a college curriculum that includes female role models, and a tradition of viewing women as active and important in the college’s history. Other strategies, such as providing a space for women and hosting campus events that focus on women’s issues, are shown to improve the academic and professional success of female students. The authors point out that such strategies are often regarded as inessential academically, and, thus, are often cut first when budgets are restricted. Combating this view, the research and statistical
A disability is an impairment substantially limiting a major life activity. Impairments are fairly easy to define, as are major life activities. However, determining the parameters of "substantially limiting" is the vexation of every disability services professional. Determining if an impairment rises to the level of "substantially limiting" rests upon the documentation provided by the student. Therefore, it is not surprising that two-thirds of the book addresses questions pertaining to the documentation of clinical conditions. Though sounding dull, if not tedious, it is in fact far more relevant and interesting than one may initially imagine. While determining the impact of traditional conditions such as hearing, vision loss, or physical disability is relatively consistent, the task is more intricate and obscure for hidden impairments. Learning disabilities, attention deficit, and mental health disorders often influence, and thus alter, the very essence of the educational process. Establishing professional standards and common conventions is essential to establishing a viable process for both accommodating students with disabilities and upholding the academic integrity of the institution. Dedicating six of the book's nine chapters to documentation issues is neither excessive nor verbose.

Michael Gordon, a professor of psychiatry and Director of the Attention Deficit and Hyperactivity Disorder Program at SUNY Health Science Center in Syracuse, NY, and Shelby Keiser, manager of the Office of Test Accommodations at the National Board of Medical Examiners, are aptly trained and prepared to edit a text on accommodations in higher education. In addition to subject matter addressed by the editors, eight other appropriately trained professionals have contributed to this well-organized and "no-nonsense" guide.

The book has been well accepted throughout higher education. Louise Russell, Director of the Student Disability Center at Harvard University states, "This book does not shy away from important questions, challenges, and controversies." Nancy C. Hill, a partner in the law firm of Carey, Hill & Scott in Charleston, West Virginia wrote that it is "The perfect primer for lawyers who want to understand how the ADA applies to higher education and professional testing." And David K. Fram, director, ADA & EEO Services at the National Employment

---

**Accommodations in Higher Education under the Americans with Disabilities Act (ADA): A No-Nonsense Guide for Clinicians, Educators, Administrators and Lawyers**

Edited by Michael Gordon and Shelby Keiser

1998 GSI Publications

236 pages; hardcover. $35.50

Readers of Accommodations in Higher Education under the ADA are likely to know more qualifying factors relating to disability issues and protocols than many disability service providers who haven’t read the book. This brave statement is based on the fact that, until recently, there has been little in the way of acceptable conventions concerning the complexities of disability accommodations in higher education. After all, the ADA is less than ten years old and case law, the driving force in developing protocols, has been slow to develop. Gordon and Keiser’s text offers the first attempt at consolidating, documenting, and solidifying a new body of literature, data, and practices based upon both common and case law. This book offers a comprehensive and accurate summary of current interpretations and applications of the ADA. It is a must read for disability service practitioners and administrators, and a desirable read for anyone wanting a firm understanding of the facts surrounding the ADA’s execution in higher education.

A disability is an impairment substantially limiting a major life activity. Impairments are fairly easy to define, as are major life activities. However, determining the parameters of "substantially limiting" is the vexation of every disability services professional. Determining if an impairment rises to the level of "substantially limiting" rests upon the documentation provided by the student. Therefore, it is not surprising that two-thirds of the book addresses questions pertaining to the documentation of clinical conditions. Though sounding dull, if not tedious, it is in fact far more relevant and interesting than one may initially imagine. While determining the impact of traditional conditions such as hearing, vision loss, or physical disability is relatively consistent, the task is more intricate and obscure for hidden impairments. Learning disabilities, attention deficit, and mental health disorders often influence, and thus alter, the very essence of the educational process. Establishing professional standards and common conventions is essential to establishing a viable process for both accommodating students with disabilities and upholding the academic integrity of the institution. Dedicating six of the book’s nine chapters to documentation issues is neither excessive nor verbose.

Michael Gordon, a professor of psychiatry and Director of the Attention Deficit and Hyperactivity Disorder Program at SUNY Health Science Center in Syracuse, NY, and Shelby Keiser, manager of the Office of Test Accommodations at the National Board of Medical Examiners, are aptly trained and prepared to edit a text on accommodations in higher education. In addition to subject matter addressed by the editors, eight other appropriately trained professionals have contributed to this well-organized and "no-nonsense" guide.

The book has been well accepted throughout higher education. Louise Russell, Director of the Student Disability Center at Harvard University states, "This book does not shy away from important questions, challenges, and controversies." Nancy C. Hill, a partner in the law firm of Carey, Hill & Scott in Charleston, West Virginia wrote that it is "The perfect primer for lawyers who want to understand how the ADA applies to higher education and professional testing." And David K. Fram, director, ADA & EEO Services at the National Employment
Law Institute states that, "This book provides well-written, practical and user-friendly guidance for professionals who deal with students claiming a need for accommodations in instruction or testing."

If strapped for time, readers will benefit greatly from reading only the first chapter in the Essential Concepts/Administrative Considerations section. Here the authors define and explore the six basic principles creating the ADA's foundation as applied to higher education.

Accommodating students with disabilities in higher education will remain a complex and enduring issue for college administrators. Appreciating the intricacies involved can only serve to improve the ability of collegiate registrars, admissions officers and other related university personnel, in meeting the needs of students as well as protecting the best interests of the university.

—J. Trey Duffy, Director
McBurney Disability Resource Center
University of Wisconsin-Madison

LETTERS TO THE EDITOR

Pakistani Educational Credentials

Dear Editor:

First I want to congratulate you on the new format of C&U. It is easy to read and very functional.

I also read with great interest "The Forum" on pages 28-30 of the Winter 1999 issue (Volume 74, No. 2), and find the answer to the question about Pakistan credential equivalency to be at variance with the reality. For example, Pakistani students matriculate completing ten years of schooling. Then they either complete three years of college to earn a B.A. or B.S., or two more years to obtain a master's degree. Firsthand knowledge of the system leads me to conclude that such a B.A. or B.S. is equivalent to one year of college whereas the said M.S. would be equivalent to three years of a four-year university curriculum in the United States.

Furthermore, the student preparation at English Media Schools/Universities with similar programs are far better than those from public institutions with Urdu as medium of instruction, and would command full English credit at the high school or university level. In summary, a course-by-course content/depth analysis is essential for granting U.S. course credit. Another recommendation is to require either an English exam or coursework to ensure an applicant's academic success.

I have over thirty-eight years of U.S. university teaching, over ten years of industrial experience worldwide, and thirty-six years of credential evaluation experience.

Dr. Harb S. Hayre
Director, C.E.I. Education Specialists, Inc.

Dear Editor:

Dr. Hayre is incorrect in stating that students in Pakistan "matriculate" upon completing ten years of education. "Matriculate" is a term that is no longer in formal academic use in Pakistan. No matriculation certificate is actually awarded these days by any university or by any other institution or agency in Pakistan. Upon completion of ten years of primary-secondary education, students in Pakistan receive a Secondary School Certificate (SSC).

Those who have received an SSC continue their education in a two-year "intermediate" program, also referred to as a "higher secondary" program, upon completion of which they receive an Intermediate Certificate (IC) or Higher Secondary Certificate (HSC). This is the credential that enables a student to actually enroll at a university.

Universities in Pakistan offer two-year programs leading to a Bachelor of Arts, Bachelor of Commerce, or Bachelor of Science degree, each of which can be followed by a two-year master's degree program. Some students complete a three year Bachelor of Arts (Honours) or Bachelor of Science (Honours) degree program instead. They can complete a one-year master's degree program. There is also a three-year bachelor's degree program in pharmacy; it leads to a two-year master's degree program. Four-year bachelor's degree programs are offered in engineering, and five-year bachelor's degree programs are offered in architecture.

Most foreign educational credential evaluators in the United States, following the advice presented in the AACRAO publication on Pakistan, consider a bachelor's degree in engineering or architecture from a university in Pakistan to be the equivalent of a bachelor's degree in the same academic field in the United States. They consider the other bachelor's degrees awarded by universities in Pakistan to merit one year of possible transfer of credit for each year of full-time academic work completed in Pakistan.

James S. Frey, Ed.D., President
Educational Credential Evaluators, Inc.
FAX or MAIL Directly to Our Advertisers—
Find Out What They Have to Offer!

Select the advertisers that interest you and fax or mail this form to them for more information.

- □ I would like more information about ____________________________.
- □ Please call me with information.
- □ Please mail me information.

*Please print, type, or attach business card*

- □ C.E.I Education Specialists
  Dr. Harb S. Hayre, Director, 10 Legend Lane,
  Houston, TX 77024-6003
  Phone (713) 952-6753; Fax (713) 467-2474

- □ Educational Testing Service–TOEFL
  (inside front cover) TOEFL Program Office,
  PO Box 6155, Princeton, NJ 08541-6155

- □ Quodata Corporation
  (inside back cover) Meg Brenneman
  One Union Place, Hartford, CT 06103;
  Phone (860) 728-6777; Fax (860) 247-0249

Name ____________________________________________ Title ________________________________
Institution/Company ________________________________________________________________
Address ____________________________________________________________
City ___________________________ State ________ Zip+4 ________________________________
Phone __________________________ Fax __________________________
E-Mail ____________________________

Spring 1999
"Promise delivered. Everything promised."

"Flexibility of the software exceeded my wildest expectations!"
Marty Eppley, Registrar and Associate Dean, Elizabethtown College.

Quodata
One Union Place
Hartford, CT 06103
860 728-6777
www.quodata.com

Information Management for Higher Education