Selection Through Individualized Review

A Report on Phase IV of the Admissions Models Project

Gretchen W. Rigol

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 4,500 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com.

Copyright © 2004 by College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. Connect to college success, SAT Reasoning Test; and SAT Subject Test are trademarks owned by the College Entrance Examination Board. PSAT/NMSQT is a registered trademark of the College Entrance Examination Board and National Merit Scholarship Corporation.

TABLE OF CONTENTS

INTRODUCTION 1

DIFFERENT APPROACHES TO
INDIVIDUALIZED REVIEW 3

STANDARDIZING ELEMENTS
OF THE APPLICATION 9

READER TRAINING AND GUIDELINES 17

CONSISTENCY AND RELIABILITY 21

THE ECONOMICS OF
INDIVIDUALIZED REVIEW 23

DEFINITIONS 27

ADDITIONAL THOUGHTS
AND NEXT STEPS 31

REFERENCES 33

APPENDIX A 35

ROSTER OF PARTICIPANTS IN THE
JUNE 8–9, 2004, CONFERENCE

APPENDIX B 39

CONSISTENCY AND RELIABILITY
BACKGROUND PAPER

INTRODUCTION

This is the fourth monograph in the Admissions Models series. It is based, in large part, on the discussions that took place at a College Board conference on June 8–9, 2004, in Chicago. That meeting differed somewhat from prior Admissions Models Conferences in that it included researchers and individuals representing professional schools and other organizations, as well as experienced undergraduate admissions professionals. A complete listing of participants is provided in Appendix A.

In preparing this report, additional information was gathered from other admissions experts and from other fields that might inform the discussion about selection through individualized review. For example, in considering different ways to train readers and evaluate an applicant's personal statement, essay-scoring procedures used in other settings were examined. Other types of selection procedures, such as those used in competitive scholarship programs and in hiring new employees, were also investigated.

The overall goal of this project was to gather as much information as possible about different approaches to selection, with the specific objective of helping colleges and universities evaluate their own practices and possibly identify ways to improve their procedures. As has been evident in the previous work on admissions models, there is no single best way to select students that would be appropriate for all institutions. And, at the same time, it is impressive and remarkable to learn of the thoughtful and thorough processes that different institutions have implemented and to see their continuing interest in seeking ways to enhance the way they select students.

Readers of this report are encouraged to read the three earlier monographs in this series. The first report, *Toward a Taxonomy of the Admissions Decision-Making Process* (1999), identifies the different philosophical approaches to selection. *Best Practices in Admissions Decision* (2002) outlines various considerations that constitute best practices in the field. *Admissions Decision-Making Models: How U.S. Institutions of Higher Education Select Undergraduate Students* (2003) examines the components of the application and the different ways they are evaluated, and summarizes the different models in use.

The title of this report is clearly inspired by the June 2003 Supreme Court decisions in the University of Michigan admissions cases. Although the term "individualized review" was not used frequently prior to the Michigan decisions, the actual practice of individualized review has been widespread at both public and private universities for many decades, if not longer. As noted in the chapter on definitions, many terms are used to describe the individualized review process—holistic, comprehensive, judgmental, and

whole-file review. While there are various approaches to this type of review, there are few generally accepted distinctions implied by the different terminology.

I am grateful to the many colleagues who contributed to this report by sharing information and taking time to explain the inner workings of their selection procedures. Special thanks go to Delsie Phillips, dean of admission and financial aid at Haverford College, for serving as chair of the June 2004 conference, and to Wayne Camara, Glenn Milewski, and Emily Shaw from the research department who helped identify resources and prepared background documents. I would also like to thank Fred Dietrich, senior vice president for higher education, for his continuing support of the Admissions Models project and for asking me to continue to be involved in this work.

Gretchen W. Rigol
Consultant
September 2004

DIFFERENT APPROACHES TO INDIVIDUALIZED REVIEW

Selection through individualized review can be accomplished in a variety of ways, ranging from a general reading of a candidate's file by one or more evaluators, which results in a single overall rating to a highly structured analysis of many different factors about each applicant. There are, however, common elements in every approach.

Depending on the institution's mission, some combination of academic and personal qualities are identified, sometimes in great detail, to aid in evaluating an application. Academic credentials are given the greatest weight in all admissions processes examined; however, there were no competitive admissions models that disregard personal factors. Although each institution defines personal factors differently, some of the more common qualities that many institutions look for are leadership, contributions to community, intellectual curiosity, special talents, life experiences, personal circumstances, and other background variables such as socioeconomic status and, in some cases, racial/ethnic status and sex.

It should be emphasized that individualized review does not mean that unqualified students are gaining admission at the expense of other more highly qualified students. The fact is that most of the applicants to competitive institutions have strong academic credentials. Colleges and universities select students to shape their incoming classes in ways that are consistent with the institutions' mission and goals. They want students with different backgrounds and experiences and different strengths and talents. That is why some of the most outstanding applicants might be accepted at some highly competitive universities, but denied at others.

In some cases, individualized review is just part of the process. Sometimes there is a "triage" approach that automatically admits some subsection of the application population based on a predetermined standard—such as the top 4 percent, the top 10 percent, or other class rank qualification. In other circumstances, there might be institutionally specific data that justifies decisions either to admit or deny an applicant based on academic credentials (GPA, class rank, and/or test results). Most experts agree that it is justifiable to make such a decision if there is empirical evidence, rooted in validity research, to sort students into decision categories provided that such sorting is done consistently and without regard to any other information in an applicant's file.

Individualized review generally means evaluating all of the available information about an applicant. Despite the comprehensiveness of many applications, there is no way that readers can know everything about an applicant. As one admissions dean put it, whole-file review means just that, not full-life review.

Although processed differently, individualized review contains both objective and subjective elements. Most institutions use some sort of rating scales, and some combine numerical ratings into a formula that generates the final decision. Other common elements in all individualized review processes include specific guidelines to assist readers in evaluating a file and reader training. (See the chapter on reader training and guidelines for more information on this topic.)

There are relatively few differences between public and private institutions. One difference is that public institutions often have defined some minimum level of academic qualification or eligibility, while private institutions are more likely to outline recommended academic preparation and expected levels of achievement. Public institutions are also somewhat more likely to have highly structured approaches to the review process.

The following five examples illustrate the range of possible ways that individualized review is organized. Each is based on actual practice, but some details have been modified to demonstrate the range of approaches that are used. And there are undoubtedly many other approaches that are possible. Although all of these examples come from institutions that receive large numbers of applications for relatively few places, prior research revealed similar approaches at less competitive colleges.[1]

Example A

This highly competitive university (with about 15,000 applications to review) assigns folders to two-person teams (comprised of one senior member of the admissions staff and one junior staff member, graduate student, or other part-time reader). Files are read according to the particular school within the university to which the student has applied. Each application is rated separately on three dimensions: the strength of the academic record; communication (based on the applicant's essays and responses to short-answer questions, and teacher and counselor comments); and character, leadership, and initiative. Guidelines suggest that the academic rating should be weighted about 60 percent, with communication and character each contributing 20 percent to the overall rating. The purpose of these "subratings" is to assure consistency across readers and to impose a common way of evaluating all applications. However, the subratings are not automatically combined to create an overall rating.

Instead, after completing the review of the entire file, the reader assigns an overall rating based on his or her evaluation of all of the information reviewed. The overall rating scale is 1–5, with 1 being the highest score. If both

1. For additional examples, together with detailed information about factors considered and approaches to evaluation, see *Admissions Decision-Making Models: How U.S. Institutions of Higher Education Select Undergraduate Students* (2003).

readers agree on the overall rating, the decision is done. 1s and 2s are admits, 3s are held until the end of the process, 4s are usually placed on the waiting list or denied, and 5s are denied. If both members of the team do not agree on the overall rating, the file goes to the dean, director, or senior associate for a final rating. There are no real committees, but teams meet weekly to discuss puzzling issues they've encountered during the prior week of reading.

Example B

Another competitive institution (with about 12,000 applications) has a somewhat similar system, but requires three complete readings of each folder. All files are assigned randomly, with the first two readings being "blind" (meaning that neither reader is aware of the first two readers evaluations.) The third reader is aware of the ratings of the earlier readers. The evaluation is highly structured and based on a nine-point scale for both academics and personal qualities. The process is individualized in that each student is considered one by one, and it is holistic in that it considers each applicant "as a whole," taking into account all academic and personal dimensions the student demonstrates through his or her application.

As in Example A, each reader gives the file an overall rating—in this case, admit, wait list, or deny. The final decision is based solely on the overall ratings of the three readers. The process is both thorough and efficient, and an emphasis is placed on training to assure fairness and consistency. There is no committee, in part because of a concern that the dynamics of a committee can be unpredictable.

Example C

At this competitive college (with about 20,000 applications), files are read randomly. Readers first evaluate the strength of the academic record, which is based on the transcript, test scores, teacher evaluations, and school recommendation. Thus, the first impression that the reader forms about the applicant is on academics. Readers go on to look at other information about the student's life experiences and other competitive factors that distinguish the applicant. All information in the file is read, including interview reports if available, and one reader prepares a summary of the entire file. The summaries are then compiled into a docket (arranged by schools within states).

Actual decisions are made by vote of the full admissions committee. Each applicant is presented to the committee by one of the readers and then discussed by the committee. The discussion focuses on what contributions the student might bring to the campus. Because most applicants have extremely strong academic qualifications, the committee discussions often focus on the

individual characteristics of the applicant, including race and socioeconomic status and other details about the student gleaned from personal statements, recommendations, and interviews.

This institution uses this committee model for decision making because of its democratic nature, and also because of the belief that any personal biases that one committee member might have are offset by others on the committee. For example, one member of the committee might always be impressed with Eagle Scouts, while another might favor classical pianists or science competition winners.

Example D

This large highly competitive university (with more than 40,000 applications) employs a multistep process involving individual reviews, which yield three different ratings that are then placed on a decision grid. One set of reviews focuses on two dimensions: personal achievement and life challenges. Based on a reading of the application, personal statements, and a summary of the academic record, readers assign ratings for both dimensions. Other readers conduct academic reviews, focusing on grades, course work, test results, and scholastic honors.

These three ratings are then combined on the decision grid. In general, applicants with exceptionally high academic ratings are accepted regardless of their ratings on the other dimensions. It is in the middle ranges of academic ratings where a personal achievement rating and/or a life challenges rating can make a difference. Borderline applications are reread to verify the ratings, since a single number could make the difference between acceptance and denial. Therefore, even though a computer can apply the numbers to arrive at a decision, there are many human reviews that provide the ratings that form the basis for the decision.

Training and quality control are essential components of this process. There is an extensive training program, and all readers must be certified. In addition, readers are constantly monitored to assure consistency. Readers also can consult a resource center if they have questions about how to interpret something on the application they are reviewing.

Example E

At another large university (which receives more than 25,000 applications annually), the process is partially objective and partially subjective. The objective part is a computer-generated academic achievement index based on the student's class rank and test scores (and an extra bonus if the student has taken more than the core course requirements). This index is based on validity research about the actual performance of prior classes and is essentially an estimated freshman GPA.

A second index, the personal achievement index, is then assigned to each applicant based on a holistic review of the entire application, including two essays. A faculty member experienced in the holistic review process used in grading Advanced Placement Program® Examinations and the writing section of the SAT Reasoning Test™ trains readers. Although academics are considered in this review, the emphasis is on personal achievement, particularly in the context of the student's background.

The two ratings are then combined on a separate decision grid for every college at the university. As with Example D, the higher the academic rating, the more likely the student is to be admitted. Lower academic indices may be offset by high personal achievement ratings. Students just below the acceptance lines on the grid are offered the opportunity to enroll in a special summer program and then matriculate in the fall if their performance is satisfactory.

STANDARDIZING ELEMENTS OF THE APPLICATION

Although standardized test scores are usually a part of a candidate's credentials, the information provided on the actual application and supplemental materials (such as transcripts, school profiles, recommendations, interview reports, and personal statements) may vary greatly in quality, accuracy, completeness, and usefulness. In order to be able to fairly evaluate applicants and compare them to one another, attempts have been made to standardize some of this information. Another motivation for standardization is to make the process easier for students, counselors, and others who provide information about the applicant.

At the same time, some institutions encourage readers to judge applicants individually, with little regard to applying a consistent standard. Each applicant brings a unique set of background characteristics to the table, many of which are intangible—but sometimes only valued and highly regarded by some. Institutions that encourage independence in their reading process are generally most likely to use multiple readings and to have options for democratic review through committee discussions.

The Application

The Common Application Form is now used by more than 250 institutions, although most also require institutionally specific essays or other supplements. Many states or systems also have a common application form used by all public (and, in some cases, private) colleges and universities in the state. While this is advantageous to the student, it does not always permit colleges to gather the depth of information they might find useful. For example, in one state, all public institutions, including community colleges, use the same basic application. Although there are some open-ended questions designed to help reviewers learn more about the applicant, vague prompts, such as "tell me about yourself and why you want to attend this institution," elicit very different and not entirely useful types of responses. Some states have solved this problem by having a basic application used by all institutions and a supplemental application required by only a few institutions that are more competitive or have special programs that require more information. In addition, some institutions have developed a supplemental questionnaire that is used for a subset of borderline students to gather information on issues such as an unexplained deviation in academic performance, why the student works so many hours, the time spent on homework, and more information about the student's high school.

In order to get similar information from all candidates, more colleges have added short-answer questions to their applications. Examples of some of these questions are:

- *Tell us about a talent, experience, contribution, or personal quality you will bring to the campus.*
- *How has your family history, culture, or environment influenced who you are?*
- *If you were president of the United States for a day, what one policy—whether it is serious or semiserious—would you implement? Why?*
- *Looking ahead to the next 5 to 10 years, what personal, social, or political issue concerns you most? Why?*
- *As a student at this university, what contributions do you expect to make to the campus community?*
- *Tell us which one of your school-related, work, or volunteer activities has been most meaningful to you, and why.*

The High School GPA

Recomputing an applicant's high school GPA is one of the most common ways application credentials are standardized. This is particularly important for institutions that use the GPA as part of an academic index in evaluating applicants. Institutions that employ reading as the only way of evaluating an application are less likely to recalculate the GPA than those who use this information in some formulaic way in their process.

Just as secondary schools use myriad approaches to calculating GPAs, so do colleges. There are two major elements in calculating or recalculating a GPA—determining which courses to include and deciding whether to give extra points for honors and/or AP® or other college-level work. Many colleges that recalculate a GPA examine each course title on the transcript and determine whether it meets their particular requirements. Some include only academic or college-prep courses; others have specific lists of qualified courses. Some will include theater, art, and religion courses; others won't.

There is considerable debate over the issue of weighting advanced courses. Some believe that students who take more demanding courses should be rewarded with extra points and that students might be encouraged to take easier courses if all courses are counted equally. On the other hand, there is concern that a student attending a school that does not provide students the opportunity to take advanced courses might be disadvantaged. Some institutions have adopted a compromise position by giving extra credit for only a fixed number of courses.

The Transcript

Transcripts come in many different forms. Most are now computer generated but, with few exceptions, transcripts are still submitted to admissions offices in paper format. Overall, there is no standardization in terms of course titles and

grading systems, although more and more states have adopted some level of standardization in format and/or course titles. Grading systems remain for the most part the prerogative of individual schools or districts. Because of this, many colleges reformat the secondary school record into a common format. For example, some cluster all courses sequentially by subject so that reviewers can easily see the progression of course work and grades in English, mathematics, science, foreign languages, social sciences, the arts, and other fields. Others simply translate the information onto a common form.

There are several national attempts to standardize transcripts, including a standard form promulgated by the National Association of Secondary School Principals (NASSP). In addition, using the Electronic Data Interchange (EDI), transcripts can be sent electronically (know as Speedy Express), making it easier to reformat the information. Several states have also developed programs for the electronic transmission of transcripts to other secondary schools, as well as to colleges. At least one large university asks students to provide a self-reported transcript on a scannable form. This information is later confirmed with the official transcript when the student arrives.

A related topic is how to evaluate the strength of the school. Although experienced admissions officers have developed considerable information about schools, including the quality of their teachers, and the rigor of the curriculum, many newer staff and external readers do not have this knowledge. Some institutions assemble detailed historical information about feeder schools and how well students from those schools have performed on campus. Others use data from the Enrollment Planning Service (EPS) or look at other available information, such as number of AP courses, number of AP Examinations, and data from state report cards. Although participants in the June meeting discussed the possibility of looking at No Child Left Behind data, there was no evidence that any institutions were actually using that information.

While most agreed that it is important to evaluate every application in the context of the opportunities available to them, there was some concern that students and parents from the very strong schools believe that they should receive more consideration because they have had a more rigorous curriculum. It was noted that these students already get a boost because they usually have better writing skills and higher test scores. Instead, it is the student who has not had so many opportunities who should get the extra consideration.

Test Results

Although test scores are often the only truly standardized piece of information in an applicant's file, these scores should not be interpreted with too much precision. All test results, by definition, have a standard error of measurement (SEM) that indicates the range within which the attained score reflects the student's true score. In other words, a student with an SAT® math score of 580

should be viewed as having a true score of between 550 and 610, given that the SEM for this part of the SAT is about 30 points. Readers should also understand the concept of the standard error of the difference (SED), which is a measure of whether the test scores from two different applicants indicate a genuine difference of ability. In order for two scores to reflect real differences in ability, the two scores must be more than 1.5 times the SED apart. For example, on the SAT Subject Test™ in Physics, the SED is 40, meaning that a student with a score of 600 and another with a score of 650 may indeed have the same level of achievement in the field of physics. Readers should understand these basic psychometric principles and caution should be exercised in making fine distinctions between students on the basis of test results.

Other data that can aid in the interpretation of test scores are national or state percentiles, for all test-takers and by subgroup. Some institutions look at percentile information by school. For example, if a student has a 480 on the writing section of the SAT Reasoning Test, this is slightly below average in a universal context, but when compared with all students from a particular school, this score might reflect that the student is in the top 20 percent compared with his or her peers on this particular measure.

Recommendations

Both the Common Application and the National Association for College Admission Counseling (NACAC) have developed a common counselor or school recommendation form. However, many institutions have institutionally specific forms. A summary of the components of a random selection of recommendation forms used by public and private institutions shows that there is much variation in the types of information counselors are asked to provide.

Rather than complete the institution's recommendation form, some counselors simply provide a school profile and separate letter of recommendation about the student. Although this information is often extremely useful and helps "round out" the data about the student, this approach means that each applicant has different information in his or her file. One counselor might primarily comment on personal qualities and the students' activities while in high school. Another might summarize teacher comments about the students' intellectual capabilities and achievements. And still another counselor might write about the student's strong interest in the college.

In an attempt to both elicit the most useful information and ease the burden on faculty who are asked to provide recommendations for students applying to graduate schools, Educational Testing Service (ETS) conducted surveys of graduate schools about what characteristics they look for in candidates. These surveys identified 20 to 30 characteristics (both cognitive and noncognitive). Based on this information, ETS developed a prototype of

an electronic standardized recommendation form (US patent application No. 10/244,072, filed September 16, 2002) that can be used to rate candidates on a variety of attributes that cluster into three independent dimensions: (1) cognitive ability, (2) motivation, and (3) ability to work with others. This timely research holds promise for making it easier for counselors and faculty to provide standardized assessments of applicants.

Interviews

Relatively few institutions actually require interviews of all applicants. The traditional interview model is to request students to come to campus for an interview with an admissions staff member. If students are unable to travel to the college, arrangements are made for interviews with alumni. In most cases, alumni interviewers are provided with background information about the characteristics of the incoming class, but there is relatively little formal training. Interview reports tend to be short unstructured narratives, sometimes with an overall rating or recommendation.

A handful of institutions have adopted a highly structured and required interview process and have taken extraordinary steps to try to standardize the process. Common features of these structured interviews include extensive training for interviewers, standard questions or topics to be discussed, and an evaluation form designed to elicit information about specific characteristics of the applicant. Described below are two examples.

Example F

At this moderately large university, (with about 4,500 applicants), interviews are scheduled at some 30 different locations around the country. Arrangements for telephone interviews are made if students are unable to travel to one of the interview sites, but most applicants participate in interview sessions at a location near their homes. Teams of alumni, faculty, and staff conduct 20-minute interviews designed to assess the student's motivation to learn, integrity and honesty, openness to differences and new ideas, and community citizenship and caring. Interviewers do not have copies of the student's application materials and are not even permitted to ask about what school the student attends in order to avoid any bias in the process and to assure that the interview captures different information from what is available from other sources.

Interviewers undergo formal training and are provided with extensive training materials. All students are asked the same basic lead questions (samples of which are provided in advance to students). There is also a list of sample probes that the interviewers may ask to draw more information about the topic being explored. Following the interview, each interviewer completes a standard evaluation form, which provides ratings on a five-point scale for each of the qualities described above. Interviewers are also encouraged to provide

additional comments on the form. The interview ratings are then added to the students' application materials and become one of many factors considered in the overall evaluation process.

Example G

A small specialized seven-year medical program leading to combined B.S. and M.D. degrees which is designed to train primary care physicians for medically underserved, inner-city communities, also requires interviews. This program receives about 600 applications, which are first reviewed by the admissions staff. After a thorough reading of all materials, primarily focusing on academic qualifications, approximately 250 applicants are selected as "finalists" to come to campus for interviews.

Faculty and student volunteers conduct three separate 30-minute interviews. The training for interviewers includes an overview of the demographics of the current pool of finalists and sample interview forms from the prior year. Each interviewer is required to ask specific questions and, at the end of the interview, completes a detailed evaluation form for the following five categories:

- *Life experience and connection to the world*
- *Approach to learning*
- *Commitment to the goals of the program*
- *Personal attributes*
- *Communication skills*

An overall rating is also provided. This information is then included in the student's file, which is then reviewed by the entire faculty admissions committee for a final decision.

Other Information

Over the years, there has been considerable interest in exploring whether there are other types of information, including measures of noncognitive factors, that might be useful in selecting a class. The College Board, as well as other similar organizations representing professional and graduate admissions, has explored a variety of supplemental questionnaires and new assessments that hold promise for enhancing the selection process. In early 2002, the College Board held a major symposium on new tools for admission to higher education. Published papers commissioned for that meeting focus on topics such as broadening predictors of college success; augmenting the SAT through assessments of analytical, practical, and creative skills; the case for noncognitive measures; rethinking admissions and placement in an era of new K–12 standards; and proficiency-based admissions.[2]

2. *Choosing Students: Higher Education Admissions Tools for the 21st Century* (Earlbaum Press, 2004).

Accuracy and Verification

Admissions officers and the public have become increasingly concerned about the accuracy of information provided by applicants. As more emphasis is placed on evaluating students' credentials in light of their life experiences, there is concern that students might attempt to exaggerate or fabricate information in hopes of convincing readers of disadvantages they have had to overcome. In addition, the growth of an independent application coaching industry and the availability of considerable materials on the Internet that can be adapted for personal statements have called into question how much on the application represents original work.

A thorough reading of the entire file, including counselor and teacher recommendations, is the traditional way to identify information that seems out of line. For example, a superbly written essay by a student with mediocre English grades, low writing section scores on the SAT, and no mention of strong communication skills in recommendations might be suspect. And while not widespread, admissions officers report occasionally seeing essays that are extremely similar—some of which have subsequently been found on the Internet. At least one institution is known to specifically ask on the application form if the student received any assistance in preparing his or her personal statement and, if so, how such assistance was reflected in the material submitted.

Some institutions have adopted formal verification procedures for some portion of their applicants. In some cases, counselors are asked for verification of specific information. In other cases, students are requested to provide additional information to substantiate the facts provided on their applications. Other institutions have included language on their application forms intended to deter students from embellishing their credentials. One institution calls the student at home if they seriously question something in the application so that parents are aware of the potential problem. In almost all cases, misrepresentation results in the application being withdrawn.

Reader Training and Guidelines

The integrity of an individualized review process relies heavily on having a cadre of readers who are well trained. All institutions that engage in individualized review offer some sort of training for the individuals who will be evaluating folders. For smaller institutions where there is relatively little staff turnover and where admissions officers do all reviews, the training might be relatively informal, with newer staff reading along with experienced admissions officers during the early decision cycle. Even after training, some institutions employ such a buddy system for the entire reading period. At the other end of the spectrum are elaborate training programs that require a week or more of hands-on training, homework, and eventual certification.

Material covered in training sessions generally includes information that describes the types of students the institution is seeking to enroll, specific qualities that are sought, information about the rating scale(s) that will be applied, and examples of files from the prior year that received each rating on the scale used. At some institutions, there is a single training period just prior to the reading season. Many institutions, however, have developed a structure for ongoing training to assure that the readers stay calibrated throughout the several months they are reviewing files. Sometimes these calibration sessions involve the entire group or a subset of readers meeting every week or so to read and evaluate the same group of folders and then discuss cases where their ratings differed.

Most institutions have developed, often with faculty advice and consent, specific definitions of the qualities that readers are to focus on. In some cases, readers evaluate the candidate on each quality separately; in other cases, these definitions provide a general framework for reading, but only one rating is issued. These two approaches are similar to the two major types of evaluation used for performance assessments: analytic scoring and holistic scoring. In analytic scoring, readers must evaluate performance on a number of specific dimensions, such as spelling, depth of vocabulary, sentence structure, organization, and originality. Each area receives a separate score and the individual scores are combined in some predetermined way that reflects the overall purpose of the assessment. In holistic scoring, the overall quality of the student's response is evaluated. Perhaps the best-known example of holistic scoring is the writing section of the SAT, where readers are trained to judge the essay as a whole. Definitions are provided for each of the six points on the score scale, but readers are trained not to analyze each element of the student's writing skill.[3]

3. For a complete description of the scoring guide being used for the writing section of the new SAT, see *A Guide to the New SAT Essay* (2004).

Reading guidelines can be relatively general or highly specific. They can also specify exactly where readers should look for information in the file about a particular quality or attribute. For example, if one of the evaluation criteria is leadership, readers might be pointed toward certain questions on the application and counselor recommendations. In evaluating the applicant's respect for intellectual, social, and cultural differences, it might be suggested that the reader look for evidence that the student has stretched beyond his or her comfort zone and has engaged in activities that require teamwork.

Other factors that can impact reading efficiency and reliability include a number of more intangible considerations—motivation of readers, background knowledge of the institution and a desire to admit students who genuinely reflect the goals of the institution, and indications that the work they are doing is "on target." Other intangibles that can help keep readers in peak form include being able to read at home or having time off from other responsibilities to spend quality time reading files. Although readers often have to make tough decisions, they need to keep in mind that their ultimate goal is to decide who should be given an offer of admission. As one admissions officer put it: "We are an admissions office, not a rejections office."

The following examples from reading guidelines of three different colleges illustrate the range of information that is provided for readers.

Example H

Intellectual curiosity and challenge

- *According to the faculty, this is the primary factor beyond academic achievement.*
- *Look for intellectual pursuits beyond what is required by the school or course. Has the student gone beyond the normal academic course load?*
- *Has the student participated in any academic summer or weekend programs?*
- *What is the student's motivation for learning? Look at recommendations, activities, and written statements for insight.*

Example I

Guidelines for evaluating short-answer questions and essays

- *Short. No effort. Inarticulate. Poor grammar.*
- *Superficial. Token effort. May be grammatically sound, but has no substance. No insight into the writer.*
- *Typical essay. Effort and sincerity evident. No masterpiece, but obvious thought put into essay. Although it may not be unique or special, the writer comes through as interesting.*
- *Very well written. Flows. Person may write about typical topics but does it better than most. Does something a little different or creative and does*

it well. Insightful. Essay enables you to get to know the applicant better.
- *Extremely well-written. Creative. Original. Memorable. You want to share it with the rest of the staff.*

Example J

Description of highest rating on extracurricular activities
- *Talents, accomplishments, and expertise have unusual depth and passion.*
- *Not simply talented, but brilliant and competitive on a national level.*
- *Will contribute significantly to the campus community.*
- *Leadership in most areas of involvement, multiple team captain, etc.*

One common element in most training programs is the use of "range finders," or sample files, that have been "normed" by experienced readers. For example, if an institution is using a single, five-point holistic rating scale, the range finders might include several examples at each point on the scale, showing how students with similar academic credentials might receive different ratings based on other information in their files. At some smaller institutions with experienced readers, these sample cases might be developed as part of the training. At larger institutions or those where outside readers are used, these "norming files" are usually developed prior to training.

The following examples illustrate just three different approaches that institutions use in training readers.

Example K

This large institution, which uses an analytic rating system, has about 70 readers, including a number of outside readers. A senior member of the admissions staff is responsible for training readers, as well as overseeing the entire reading process. There is a weeklong formal training period before reading begins. New readers must attend the entire training program; experienced readers receive all of the same information, but on a compressed schedule. Some of the major exercises and topics covered include:
- *An introduction of the institution's missions and goals*
- *A detailed description of the characteristics of last year's class*
- *National and state norms for test scores and other demographic data (i.e., who is in the "pipeline")*
- *A detailed review of geographical areas the institution serves*
- *Instructions on how to read a school profile*
- *Readers complete the admissions application as they would have when applying to college*
- *Readers review sample transcripts and recalculate GPAs according to the institution's specifications*
- *A detailed review of each of the evaluation categories*
- *Special training from writing experts in what to look for in student essays*

- *Samples of files that represent each score level in each evaluation category*
- *A group review and discussion of sample cases*
- *An individual review of sample cases*

Then, once a week during the entire process, groups of readers meet to review and discuss files with their team leaders. In addition, there are individual review meetings with the senior staff member responsible for the evaluation process.

Example L

At this medium-sized university, which uses both analytic and holistic scoring, reading is done by 10 experienced admissions staff and five part-time outside readers, most of whom are former admissions professionals and have been readers in prior years. Despite the fact that the readers are highly experienced, there is mandatory training for everyone during the annual staff retreat. Training includes a review of detailed class profiles from the past five years, a review of scoring guidelines for both the analytic categories and the overall rating scale, and a review of sample files representing the range of applications received the prior year. This training occurs before the fall school-visiting season in order to assure that admissions staff recruit the types of students the institution wishes to admit.

Example M

This large institution (which uses as many as 135 readers) has a four-step training process. First, all admissions staff and prospective outside readers receive written training materials that they must read (or reread in the case of returning staff and readers). Second, all readers must attend a three-hour overview of the process. Third, readers are given 20 files to read and score as homework. Finally, there is a group “norming session” (with 12–15 people at each table). Three different sets of files are reviewed at these sessions. If a reader rates all files appropriately during the first set, they are “certified.” Readers may continue with two additional sets until they are either certified or disqualified from reading.

One element of a fair selection process is that an applicant receives the same consideration and is subjected to the same level of expectation regardless of who reads and evaluates the file. There is agreement that the training process is the first and most important element in assuring consistency. Another important contributor to consistency is to have sufficiently clear and detailed scoring guidelines or rubrics. As background for this topic at the June meeting, a paper on consistency and reliability was prepared and is reproduced in Appendix B.[4] This paper covers topics such as the different types of inter-rater reliability, rater severity, components of rater training that improve reliability, and scoring rubrics.

Some institutions monitor inter-reader reliability on a routine basis. That is, they analyze the ratings that different readers have given the same applicants and calculate the number of times that there is exact or close agreement. For example, at one large institution, weekly reports are prepared for each reader, including the number of files read, the number of times that reader agreed with a second reader, and the number of readings that resulted in a third review (when ratings were more than one point apart). At least one institution helps assure that new readers are applying the guidelines properly by having an experienced reader do “shadow” readings during the beginning of the process. Most institutions, particularly those with intensive training programs, find that the agreement among readers ranges from 90 to 97 percent. If there are particular readers who are frequently out of sync with others, additional training is provided.

In addition, some institutions also measure rater severity—an indication of how hard or lenient a reader might be. This is particularly important if all or part of a folder will be evaluated by only one individual. Again, if there are wide differences among the severity of readers, additional training might be warranted.

In many cases, consistency is maintained by having at least two readers, with a third or even more readers brought in if the original ratings are not within 0.5 or 1.0 points. A few institutions help assure consistency by having one individual, often the dean or other senior staff member, review a summary card of all applicants. In other cases, there are experienced team leaders who review and confirm all final decisions. Another approach to assure reliability involves the recycling of files randomly throughout the process, or sometimes to the same reader, to see if the ratings are the same the second time through.

4. During the summer of 2004, this paper was expanded and will be published as College Board RN-20. See *Consistency and Reliability in the Individualized Review of College Applicants* by Emily J. Shaw and Glenn B. Milewski. Release date: mid-late October 2004.

It was noted that reliability could often be improved if additional information is gathered about the candidate. For example, in employment settings, adding an essay rather than increasing the number of individuals who review or interview an applicant often increases reliability.

The concept of reliability may be less useful in the selection process than that of validity defined broadly. The ultimate test of the accuracy of decisions should be that they yield a class that succeeds, not simply by getting good grades and graduating, but also whether the graduates do what is expected of them after they leave campus. If the mission of the institution is to produce leaders for the country in government, business, and/or the military, do graduates actually fulfill those goals?

Another way of judging the overall appropriateness of an admissions decision is to go back to the philosophical objectives of the admissions process.[5] If one of the underlying institutional objectives is to reward students with certain personal qualities (such as community service) or accomplishments (such as having overcome adversity), then the outcomes of admissions decision making should be judged, at least in part, against those criteria. Other desired outcomes might be to enroll students who actually contribute to the campus community or those who add enlivening effects or different viewpoints to classroom discussions.

5. See *Toward a Taxonomy of the Admissions Decision-Making Process* (1999) for a discussion of nine philosophical models.

THE ECONOMICS OF INDIVIDUALIZED REVIEW

Individualized review is expensive. The primary costs relate to the actual reading process, which is generally a function of the number of applications received, the amount of material in each folder, and the type of distinctions that readers are required to make. In moderately competitive situations, the reading can be relatively quick, with the review intended primarily to determine whether the student can handle the work. However, at institutions with a large and highly qualified applicant pool, readers must make fine distinctions among applicants—a process that takes considerable time. There are also considerable clerical costs associated with assembling and managing each applicant's file.

The majority of readers are members of the professional admissions staff. For institutions with an early decision plan and a fixed deadline for regular admissions, most staff devote one month in the fall and about three months in the spring to reading files. Thus, approximately one-fourth of admissions salaries can be assigned to the cost of reading files. At a hypothetical "average" college with 13 admissions staff[6] earning average salaries,[7] the cost of reading is about $140,000. Assuming an average of five clerical staff spending at least one-half of their time on assembling files, the cost rises to over $200,000. Further assuming an average number of applications (3,400[8]) to review, the average cost per file is $59 per applicant. One large competitive institution has calculated that it costs $109 to individually review each applicant.

It should be noted that the size of admissions offices and the number of applications received vary widely. For institutions with more than 20,000 enrolled students, the average size of the staff is 35.[9] Most of these very large institutions receive as many as 40,000 applications. Also, at institutions with highly experienced staff, the average salaries are undoubtedly higher than the average.

Although it is impossible to assign a precise figure to the actual cost of individualized review, it seems reasonable to estimate that actual costs are in the neighborhood of $50 to more than $100 per applicant. Given that most application fees range from $35 to $60, it seems reasonable to conclude that a significant part of the cost of evaluating applications is borne directly by institutional budgets.

The actual amount of time spent on reviewing applications varies widely. The norm for a file with one or more essays and school or teacher

6. From the *NACAC 2003–2004 State of College Admission Report* (p. 68).
7. From the *Chronicle of Higher Education Almanac Issue*, 2004–05 (pp. 24, 26).
8. From the College Board's *Annual Survey of Colleges 2003–2004*. In 2003–04, there were 5,398,529 applications filed at 1,595 institutions for an average of 3,385 per institution.
9. From the *NACAC 2003–2004 State of College Admission Report* (p. 68).

recommendations is 15–20 minutes. However, if the reader is expected to make comprehensive comments or summarize the file for a committee "docket," the time spent per reading can increase to 30 minutes or more. Some institutions have adopted reading processes that require at least one comprehensive review and one or more secondary confirmative reviews. The actual time spent on these secondary reviews can be relatively brief.

Another factor in calculating the length of time required to read a file is the experience of the reader. Most institutions report that all readers take more time at the beginning of the reading cycle. As the reading period progresses, readers find a natural rhythm and can read more quickly. Newer readers generally take more time on files than more experienced readers. In part, this is because experienced readers are extremely familiar with all components of the application, and they know where to look for specific information and can quickly identify unusual or outstanding factors.

Many institutions hire outside readers to supplement the admissions staff, particularly during peak reading periods. Outside readers are paid anywhere from $12–$30 per hour. The norm in 2004 appears to be about $25 per hour. Assuming a reader can read four files per hour, the cost per application for outside readers is slightly more than $6 per application. Obviously, the use of outside readers is more efficient than using admissions staff, but most institutions believe that it is important to have experienced and professional staff conduct most of the evaluations.

Soon after the Supreme Court announced its decisions in the University of Michigan cases, the admissions office at the university began making plans to adopt an individualized review process for their entire applicant pool of 25,000. Because of the scope of the change and the extremely short time to implement it, the overall cost was estimated to be approximately $1.4 million. Once the process becomes more routine, it is estimated that the additional annual costs will be approximately $1 million compared with the previous process, which provided individualized review for all applicants, but in a mechanistic, non-holistic manner.

One institution (with about 15,000 applications) that currently reads only about one-half of the total applicant pool estimated the cost to move to 100 percent individualized review. This particular institution estimates that readers read at a rate of five files per hour (or 12 minutes per file). Using graduate student readers (at a rate of $18 per hour plus a tuition waiver for one term) and two additional admissions counselors, the total additional cost would be nearly $300,000, or an additional $20 per applicant.

The actual amount that is spent on individualized review can be influenced by several factors. Perhaps most important is the amount of material that must be read. Even one additional short-answer question adds to the amount of time it takes to evaluate a file. Long essays and personal statements are extremely time-consuming. (Some colleges have reported as much as a

10 percent decline in applications as a result of adding an essay or other complexity to their applications, so the increase in time spent on a file might be more than offset by fewer numbers of applications to read. Another consideration is that even though the total number of applications might be reduced by the addition of more essay questions or other things that make the applications a bit more difficult, these changes often result in increased yield. In other words, only students who are genuinely interested in enrolling are likely to complete the more complex admissions applications.) But given the annual increases in the number of applications that students file, it is likely that most institutions can expect to see the number of applications rise over time.

Therefore, one possible way to increase efficiency in the reading process is to be sure that all of the information on the application is truly meaningful and helpful in making decisions. Another way that institutions are exploring to improve efficiency is their "back office" operations. Encouraging or even requiring online applications can help, as can imaging of all application materials.

Although costs are high, many institutions believe that the individualized review is the best way to select a class and is worth the additional expense. As a result of the Supreme Court decisions, any institution that wants to use race and ethnicity as one of the many factors they consider must employ some type of individualized review. It is testimony to institutions' strong commitment to diversity that so many have devoted the resources to provide 100 percent individualized reviews for all applicants. But even at institutions that do not consider race or ethnicity, individualized reviews are often used in order to bring in a class with varied background circumstances and academic and personal strengths. Although there is little formal research on the effects of selecting a class through individualized review, anecdotal reports suggest that faculty are more satisfied with incoming classes and counselors and other school officials feel more comfortable with the admissions decisions related to their students.

A number of terms are used, often interchangeably, to describe the evaluation of different aspects of a student's application and other credentials as part of a selective admissions process. Because the Supreme Court used the term "individualized consideration," this description of the admissions process has become widespread. Drawing on the terminology used for one approach to scoring essays and other performance tasks, other institutions describe their admissions process as holistic. Still other terms used are comprehensive, whole-file, whole-folder, and judgmental review.

The Supreme Court's understanding of what constitutes individualized review is found in the following excerpts. In *Grutter v. Bollinger*, Justice O'Connor wrote:

> *When using race as a "plus" factor in university admissions, a university's admissions program must remain flexible enough to ensure that each applicant is evaluated as an individual and not in a way that makes an applicant's race or ethnicity the defining feature of his or her application. The importance of this individualized consideration in the context of a race-conscious admissions program is paramount.*
>
> *…[T]he Law School engages in a highly individualized, holistic review of each applicant's file, giving serious consideration to all the ways an applicant might contribute to a diverse educational environment. The Law School affords this individualized consideration to applicants of all races. There is no policy, either de jure or de facto, of automatic acceptance or rejection based on any single "soft" variable.*
>
> *The Law School does not, however, limit in any way the broad range of qualities and experiences that may be considered valuable contributions to student body diversity. . . .The Law School seriously considers each "applicant's promise of making a notable contribution to the class by way of a particular strength, attainment, or characteristic—e.g., an unusual intellectual achievement, employment experience, nonacademic performance, or personal background.". . . All applicants have the opportunity to highlight their own potential diversity contributions through the submission of a personal statement, letters of recommendation, and an essay describing the ways in which the applicant will contribute to the life and diversity of the Law School.*

In *University of California Regents v. Bakke*, Justice Powell lauded the Harvard College admissions program:

> *In such an admissions program, race or ethnic background may be deemed a "plus" in a particular applicant's file, yet it does not insulate the individual from comparison with all other candidates for the available seats. The file of a particular black applicant may be examined for his potential contribution to diversity without the factor of race being decisive when compared, for example, with that of an applicant identified as an Italian-American if the latter is thought to exhibit qualities more likely to promote beneficial educational pluralism. Such qualities could include exceptional personal talents, unique work or service experience, leadership potential, maturity, demonstrated compassion, a history of overcoming disadvantage, ability to communicate with the poor, or other qualifications deemed important. In short, an admissions program operated in this way is flexible enough to consider all pertinent elements of diversity in light of the particular qualifications of each applicant, and to place them on the same footing for consideration, although not necessarily according them the same weight. Indeed, the weight attributed to a particular quality may vary from year to year depending upon the "mix" both of the student body and the applicants for the incoming class.*
>
> *This kind of program treats each applicant as an individual in the admissions process. The applicant who loses out on the last available seat to another candidate receiving a "plus" on the basis of ethnic background will not have been foreclosed from all consideration for that seat simply because he was not the right color or had the wrong surname. It would mean only that his combined qualifications, which may have included similar nonobjective factors, did not outweigh those of the other applicant. …*

The University of California system adopted the term "comprehensive review" to describe its admissions process. Comprehensive review is defined as the "process by which students applying to UC campuses are evaluated for admission using multiple measures of achievement and promise while considering the context in which each student has demonstrated academic accomplishment."[10] However, within this overall framework, individual campuses have designed autonomous and campus-specific processes to implement comprehensive review. There are three general approaches used, however, there are different ways each of these approaches can be implemented.[11]

10. *Guidelines for Implementation of University Policy on Undergraduate Admissions* (University of California, Issued 2001).
11. See "Guidelines for Implementation of University Policy on Undergraduate Admissions" from the *Eligibility and Admissions Study Group: Final Report to the President,* Appendix C (April 2004).

- *Unitary:* All applications are given a single comprehensive score by two readers who consider academic performance in context of school attended, family income and parent's occupation, education level, and students' personal circumstances. Admissions decisions are made based on the linear ranking of students' read scores.
- *Fixed weight:* Academic and nonacademic factors are assigned a predetermined number of points. Academic factors account for approximately 75 percent of the overall points. Admissions decisions are based on a linear ranking of students' combined scores.
- *Matrix:* Applicants are assigned points for a number of attributes. Admissions decisions are based on where a student falls on a two- or three-dimensional matrix.

Another way to differentiate different review processes is to consider how readers actually evaluate an application. As noted in the section on reader training and guidelines, there are essentially two basic approaches to evaluation—holistic and analytic.

In a traditional holistic evaluation process, information about all aspects of the candidate is considered together as a whole. This process is widespread in scoring of essays and other performance-based assessments. The following description of holistic scoring of the new SAT essay closely resembles what is intended in holistic review in the admissions context.

> *In holistic scoring, a piece of writing is considered as a total work, the whole of which is greater than the sum of its parts. . . . A reader does not judge a work based on its separate traits, but rather on the total impression it creates. Holistic scoring recognizes that the real merit of a piece of writing cannot be determined by merely adding together the values assigned to . . . separate factors . . . It is how these separate factors blend into and become the whole that is important. Holistic scoring evaluates this whole equitably and reliably.*[12]

Although the term "analytic review" is not commonly used in admissions, it appears that many institutions use more of an analytic approach than a holistic approach to evaluating applications. In testing and measurement, analytic scoring is defined as:

> *Analytic Scoring: A method of scoring in which each critical dimension of performance is judged and scored separately, and the resulting values are combined for an overall score. In some instances, scores on the separate dimensions may also be used in interpreting performance.*[13]

12. From *A Guide to the New SAT Essay* (2004).
13. *The Standards for Educational and Psychological Testing* (1999).

As has been illustrated above, many institutions combine more than one of these approaches in making admissions decisions. Some may use a holistic process for final decisions, but they may include a conscious evaluation of several specific dimensions (as in "analytic review") to assure consistency and to improve reliability. Institutions have developed many hybrid systems over the years, and there is no single best approach that is appropriate for all institutions. The ultimate tests are whether the process is fair, equitable, and consistent and if the methods used in evaluation are reliable and valid.

One experienced admissions dean noted that these relatively new terms had confused the public and could constrain the profession. What in the past was simply recognized and accepted as "admissions review" had somehow become more complex. The bottom line remains: making difficult choices and deciding which students, from a larger pool of very qualified applicants, should be admitted. And the unfortunate outcome of any competitive admissions process is that some students and their families and schools will be disappointed with the decisions.

One of the major themes that permeated the discussion during the June 2004 conference was the increasing public interest in transparency and accountability in the admissions process. This is of particular concern to public institutions, but all institutions noted the increasing pressure from constituents to explain why certain decisions were made. Most institutions have made concerted efforts to describe to the public the criteria they consider and the processes they use to make decisions but, nonetheless, there remains skepticism and concern about the decision-making process. The College Board was urged to use whatever influence it has to try to better inform the public about the considerable care and thoroughness that institutions employ in deciding who is and who is not offered admission.

A significant problem facing institutions with competitive admissions is that so many of their applicants have very strong academic records, high test scores, and many other desirable attributes (extracurricular activities, leadership, and stellar personal qualities). All applicants cannot be admitted, so from an individual perspective, students and parents question the process. At the same time, there is recognition that there are very capable students who may not be admitted to any of the institutions they have chosen as their top choices. Part of this problem may be attributed to grade inflation, to the perception that recentered SAT scores mean something more than they really do (particularly a sense of fine distinction of precision at the top of the scale), or to a sense of entitlement. Regardless of the reason, the admissions community is concerned that there are students and families that feel they have been somehow misled or misserved by the process.

Another question raised at the June meeting was whether individualized review actually improved the quality of decisions. Institutions that have long used some type of individualized review believe strongly in its appropriateness and validity. Institutions that have only recently adopted such procedures have an opportunity to study whether, and if, their decisions have yielded a better class than might have been admitted under a more formulaic and less-individualized review process. Some have hypothesized that the long-term positive effects of individualized review might be noticed in other areas, such as students challenging themselves more since they can't anticipate automatic acceptance based on grades and test scores alone.

But the ultimate challenge is to ensure that all of the students in schools today become fully prepared for college and to help students and families understand the range of very good institutions that are available to them. It is in part a pipeline problem and in part a problem of the public's narrow view of the available options for higher education. Too much of the attention about admissions in recent years has been focused on highly competitive and well-known institutions. From a national perspective, more

students from all backgrounds need to have the opportunity to develop the solid credentials that will put them in the running for the most competitive admissions processes. And all students and their families need to remember that there are more than 3,600 different institutions to choose from.

REFERENCES

Camara, Wayne J., and Ernest W. Kimmel (ed). *Choosing Students: Higher Education Admissions Tools for the 21st Century*. Mahwah, NJ: Earlbaum Press, 2004.

Chronicle of Higher Education. *Chronicle of Higher Education Almanac Issue 2004–2005.* Washington, DC: The Chronicle of Higher Education, Inc., 2004.

College Board. *A Guide to the New SAT Essay.* New York: College Entrance Examination Board, 2004.

College Board. *Annual Survey of Colleges 2003–2004*. New York: College Entrance Examination Board, 2003.

College Board. *Best Practices in Admissions Decision*. New York: College Entrance Examination Board, 2002.

Hawkins, David A. *NACAC 2003–2004 State of College Admission Report.* 2003–2004 Alexandria: National Association for College Admission Counseling, 2004.

The Joint Committee on Standards for Educational and Psychological testing. *The Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association, 1999.

Perfetto, Greg, (Primary Author) *Toward a Taxonomy of the Admissions Decision-Making Process.* New York: College Entrance Examination Board, 1999.

Rigol, Gretchen. *Admissions Decision-Making Models: How U.S. Institutions of Higher Education Select Undergraduate Students*. New York: College Entrance Examination Board, 2003.

Shaw, Emily J. & Milewski, Glenn B. *Consistency and reliability in the individualized review of college applicants.* New York: College Entrance Examination Board, 2004.

University of California. *Eligibility and Admissions Study Group: Final Report to the President.* Oakland: University of California, 2004.

University of California. Guidelines for Implementation of University Policy on Undergraduate Admissions, 2001.

Roster of Participants

Conference on Admissions Models—Phase IV
Selection Through Individualized Review
June 8–9, 2004
The Westin at O'Hare Airport

John Barnhill, Director of Admissions and Records, Florida State University, Tallahassee, FL

Michael Behnke, Vice President for Enrollment, University of Chicago, Chicago, IL

Pam Burnett, Director of Admissions, University of California, Berkeley, Berkeley, CA

Wayne Camara, Vice President, Research and Psychometrics, College Board, New York, NY

Arthur L. Coleman, Partner, Holland & Knight LLP, Washington, DC

Fred Dietrich, Senior Vice President, Higher Education, College Board, New York, NY

Steve Farmer, Senior Associate Director of Admissions, University of North Carolina, Chapel Hill, NC

Daniel Fraley, Student Service Regional Manager, Institute of International Education—Midwest Regional Center, Chicago, IL

Mabel Freeman, Assistant Vice President, Undergraduate Admissions, Ohio State University, Columbus, OH

Karl Furstenberg, Dean of Admissions and Financial Aid, Dartmouth College, Hanover, NH

Barbara Gill, Director of Undergraduate Admissions, University of Maryland, College Park, MD

Mike Griffin, Assistant Vice Chancellor, University of Denver, Denver, CO

Don Hossler, Vice Chancellor for Enrollment Services, Indiana University, Bloomington, IN

Michael Jones, Director of Admissions, U.S. Military Academy, West Point, NY

Ellen Julian, Assistant Vice President and Director, MCAT, Association of American Medical Colleges, Washington, DC

Patrick Kyllonen, Research Center Director, Educational Testing Service, Princeton, NJ

Maria Laskaris, Director of Admissions, Dartmouth College, Hanover, NH

John Latting, Director of Admissions, Johns Hopkins University, Baltimore, MD

Jerry Lucido, Vice Provost for Enrollment Management and Director of Admissions, University of North Carolina, Chapel Hill, NC

Chris Lucier, Associate Director, Undergraduate Admissions, University of Michigan, Ann Arbor, MI

Glenn Milewski, Assistant Research Scientist, College Board, New York, NY

Greg Perfetto, Associate Provost, Vanderbilt University, Nashville, TN

Delsie Phillips, Dean of Admission and Financial Aid, Haverford College, Haverford, PA

Bruce Poch, Vice President and Dean of Admissions, Pomona College, Claremont, CA

Greg Pyke, Senior Associate Dean of Admissions, Wesleyan University, Middletown, CT

Brad Quin, Executive Director, Admission Services, College Board, Reston, VA

Gretchen Rigol, Consultant, New York, NY

Erica Sanders, Director of Training, Review, and Evaluations, Admissions, University of Michigan, Ann Arbor, MI

Stephanie Sanders, Associate Director for Competitive Admissions Administration, Ohio State University, Columbus, OH

Susan Santana, Associate, Holland & Knight LLP, Washington, DC

Paul Seegert, Assistant Director for Counseling and Outreach, University of Washington, Seattle, WA

Rob Seltzer, Director of Admissions, University of Wisconsin, Madison, Madison, WI

Rick Shaw, Dean of Admissions and Financial Aid, Yale University, New Haven, CT

Bob Spatig, Senior Associate Director of Admissions for Marketing, Recruitment, and Selection, University of Georgia, Athens, GA

Keith Todd, Director of Undergraduate Admissions, Northwestern University, Evanston, IL

Vu Tran, Director of Undergraduate Admissions and Relations with Schools, University of California, Los Angeles, CA

Jim Vaseleck, Associate Counsel and Assistant to the President, Law School Admission Council, Newtown, PA

Richard Vos, Vice President, Dean of Admission and Financial Aid, Claremont McKenna College, Claremont, CA

Bruce Walker, Vice Provost, Director of Admissions, University of Texas, Austin, TX

Mark Weichold, Dean of Undergraduate Programs, Texas A&M University, College Station, TX

Sue Wilbur, Director of Undergraduate Admissions, University of California, Oakland, CA

Gita Wilder, Social Research Scientist, Law School Admission Council, Newtown, PA

Consistency and Reliability

Reliability refers to the consistency and stability of a measure from one use to the next. An unreliable measure contains measurement error. For example, if you got on your bathroom scale and it read 145 pounds, you got off and then on again and it read 139 pounds, and you repeated this process and it read 148 pounds, your scale would not be very reliable and would contain measurement error. If, however, in a series of weighings you got the same answer (145 pounds), your scale would be reliable—even if it was not accurate and you really weighed 120 pounds.

Inter-rater reliability refers to agreement among raters, or the extent to which raters judge phenomena in the same way. There are several aspects of inter-rater reliability. The first is observer agreement; it can be evaluated by correlating ratings based on observations made by different judges on the same group of people. The second is rater consistency; it can be evaluated by calculating the percent agreement between different ratings on the same group of people. A third aspect of inter-rater reliability is inter-rater severity that involves the degree of leniency or stringency of different judges; this aspect of reliability can be evaluated by comparing average ratings between different raters. Each aspect of inter-rater reliability is important to evaluate.

Evaluation of inter-rater reliability requires scores from two or more independent raters on an appropriately selected sample of students. Observed agreement can be evaluated by correlating one reviewer's scores with another reviewer's using a statistical calculation that produces a number ranging from –1 to 1, whereby 1 is equivalent to a perfect positive correlation and 0 indicates a complete lack of agreement between reviewers. Rater consistency can be examined by calculating the proportion of times that applicants' admissions materials receive exactly the same scores from a pair of raters and/or the proportion of scores that fall within ±1 point of each other. For example, imagine that two readers are given 100 applications to rate on a three-category checklist. The first category is for applications that are "not qualified," the second category is for applications that are "questionably qualified," and the third category is for applications that are "definitely qualified." If the two readers checked the same category for 90 of those applications, then the percent of agreement between readers would be 90 percent. Another valuable calculation, particularly for instances when an application is examined and rated by only one reader and this reader will not be assessing every application, is to find the average scores assigned by different readers for either the same applications or for all applications they have assessed over time. These averages

would reflect inter-rater severities. It can provide insight into whether some readers are more lenient or stringent in their scoring, and the possible need for further discussion on scoring or "calibrating" the readers.

For the purposes of individualized review in the college admissions process, reliability becomes a major concern when a number of different reviewers evaluate and make important recommendations or actual decisions based on somewhat subjective application materials. In this instance, the concern is not as much with the reliability of the applicant's essay, SAT scores, or high school grades over time; instead, it is with the reliability of the rater or reader. In other words, the focus is on the consistency of ratings of admissions materials between two or more readers or by different readers in settings where only one reviewer rates an application. When only one reader reviews a file and different readers are responsible for a certain number of files, the concern arises about whether or not some readers may be more lenient or stringent than others when making judgments about the applicants' qualifications. If reader ratings or decisions are unreliable, it is likely that when another reader reviews an application, this new reader's rating and decision would be different from a previous rating or decision.

It is important to note that some variation in scores for an individual on a particular measure is expected since no examinee or rater is completely consistent. However, this variation should not be unduly influenced by measurement error. Sources of measurement error can be thought of as either internal or external to the examinee. Internal sources of measurement error may include the person's level of motivation, interest, attention span, and amount of fatigue, or health, which can affect the neatness, completeness, or level of detail of the application materials. Measurement error that is external to the examinee may include scorer subjectivity or variation in scorer standards.

Effects of scorer subjectivity and variation in scorer standards both play a role in inter-rater reliability. Evaluating inter-rater consistency is highly important to any assessment that must be judgmentally scored. In an individualized selection process, not only are the applicants receiving judgmental scores or ratings, but also these scores become part of important decisions. It should be noted, however, that inter-rater reliability does not take into account the consistency of an individual's scores across different tasks, or how consistently an individual performs or scores on the essay, the academic transcript, the activities or community involvement, or the interview in the review of that individual's application. This type of reliability is most similar to internal consistency reliability, which is used to determine the consistency of results across items (e.g., essay, transcript, awards) within a test (which can be thought of as the entire application).

It is important to compare the average scores of each reader in order to check whether there is a strong tendency for one reader to be consistently

more or less lenient than another. Such a situation might result in some students receiving higher scores because a reader is more lax and some students receiving a lower score because a reader has higher expectations. One helpful way to encourage reliability between readers is to have the readers meet somewhat regularly to discuss their ratings of several of the same applicants and their reasoning behind the scores they assigned. If disagreements arise, the readers can discuss them and arrive at rules or guidelines for assigning particular scores for particular measures (essays, letters of recommendation, academic transcripts).

Rater training is a good method to improve reliability and reduce measurement error, especially for assessment procedures that require subjective judgments to be made on constructed responses. This training might require the participation of admissions counselors, professors, alumni, or anyone who is involved in the individualized review process. The training might focus on informational and/or practice sessions aimed at identifying and agreeing upon the constructs that are being assessed in the individualized review, as well as how these constructs can/should be most appropriately measured. Training has been found to increase rater self-consistency, though it is not necessarily the most effective way of eliminating differences in rater severity, or amount of leniency or stringency in scoring. Training seems to bring the extreme scorers within a more tolerable range of severity, but it will not eliminate differences in reader severity. Major differences in severity that arise in the individualized review process will likely require significant dialogue between all readers involved as this may be the result of differing definitions of the ability that the assessment is intended to measure and score.

Rubrics are another way to improve inter-rater reliability. Rubrics facilitate rater agreement by explicitly outlining the standards or achievements that correspond to different ratings. A rubric usually consists of ordered categories coupled with descriptions of criteria that match these categories, which assist in assigning levels of achievement to student-produced material. Scoring criteria in rubrics should reflect the content and processes judged by the admissions committee to be important. Creating well-defined, detailed rubrics requires the college or university admissions committee to make clear value judgments and determine the most important and critical aspects of performance/achievement/potential that the school is looking for in an applicant. For example, a rubric used to review an individual's extracurricular activities, service, and leadership may include categories, such as awards and honors received, community service, leadership positions, etc., to be rated on a 1–10 scale. This can help to "standardize" the process, enhancing consistency, as well as explicitly defining what is important to the institution. Therefore, the decision of what to include in a rubric for the individualized review process should be deliberate and well thought-out.

The new SAT is an example of a situation where subjective ratings will be made reliable because of rigorous reader training and the use of a detailed rubric. The new SAT, which will be administered for the first time in March 2005, includes an essay that will be scored independently by two qualified readers. Essay readers will be required to participate in an online training course that familiarizes them with the principles of holistic scoring and teaches them to evaluate essays according to the scoring rubric developed by the College Board. The rubric includes detailed criteria, structured on a six-point scale, of the qualities that distinguish an essay at each scoring level.

Because of the rigorous training and high qualifications of the readers, combined with the detail of the rubric, the College Board expects that more than 92 percent of all scored essays will receive ratings within ±1 point of each other on the six-point scale. If the two readers' scores differ by more than one point, a third reader will score the essay. The third reader will always be a highly experienced, veteran reader who likely provides the training to other readers on how to score the essay, and will assign a score to the essay that will become the person's final rating. Another frequent method of score resolution involving three ratings or scores by readers is to average the expert's score with the score of the rater that is closest to the expert's score. An additional way to compensate for rater inconsistency on other measures may be to have two or more raters independently score the measure and use the average of the two or more ratings. Better reliabilities can be obtained by using more raters.