2014-06-23

Are Undergraduate GPA and General GRE Percentiles Valid Predictors of Student Performance in an Engineering Graduate Program?

Larry L. Howell  
*Brigham Young University - Provo*, lhowell@byu.edu

Carl D. Sorenson  
*Brigham Young University - Provo*, sorensen@byu.edu

See next page for additional authors

Follow this and additional works at: [https://scholarsarchive.byu.edu/facpub](https://scholarsarchive.byu.edu/facpub)

Part of the [Mechanical Engineering Commons](https://scholarsarchive.byu.edu/facpub)

**Original Publication Citation**  

**BYU ScholarsArchive Citation**  
[https://scholarsarchive.byu.edu/facpub/1343](https://scholarsarchive.byu.edu/facpub/1343)

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen amatangelo@byu.edu.
Are Undergraduate GPA and General GRE Percentiles Valid Predictors of Student Performance in an Engineering Graduate Program?*

LARRY L. HOWELL, CARL D. SORENSEN and MATTHEW R. JONES
Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602, USA. E-mail: lhowell@byu.edu, sorensen@byu.edu, mrjones@byu.edu

While both subjective measures and quantitative metrics play an important role in admissions decisions, quantitative metrics are amenable to critical analysis using the tools of academic analytics. The hypotheses that motivated this study are: 1. Can an applicant’s undergraduate grade point average (UGPA) and scores on the Graduate Records Examinations (GRE) be used to accurately predict the performance of the applicant in a graduate mechanical engineering program? 2. Is a single construct based on these quantitative predictive metrics a valuable tool in efficiently making admissions decisions? This study analyzed the relationship between quantitative predictive metrics, available at the time of application to a mechanical engineering graduate program, and quantitative performance assessments measured at the thesis defense. The sample includes 92 students graduating with MS degrees in mechanical engineering from a private university in the United States. The input variables include UGPA, and percentiles for the verbal, quantitative, and written sections of the GRE. The performance metrics were obtained at the thesis defense. They are graduate grade point average, months to graduation, peer-review publication rating, and advisor determined performance rating. Each variable was normalized and the relationship between the predictive metrics and the performance metrics was analyzed statistically. Regression models were created for each performance metric and for a weighted sum of all the performance metrics. The dominant predictors are the UGPA and the score on the quantitative section of the GRE. A quantitative application rating is found to be 5 times the normalized UGPA plus four times the normalized score on the quantitative section of the GRE. Quantitative metrics account for one fifth of the variance in the performance metrics. The Quantitative Application Rating—a single construct based on the quantitative predictive metrics studied—aids in making admissions decisions.

Keywords: graduate admissions; admissions validation study; academic analytics

1. Introduction

The significant investment made in higher education is justified by the fact that a well-educated workforce strengthens economies and promotes stability and security. The need to keep pace with the rapidly expanding knowledge base requires continuous evaluation and improvement of the policies and procedures implemented in higher educational systems. However, there is mounting evidence that the quality of learning is declining in the U.S. and that many new graduates are inadequately prepared to compete in a global, knowledge-driven economy [1] in spite of efforts to improve. The recent publication of Is College Worth It? by former U.S. Secretary of Education William J. Bennett [2] has highlighted these growing concerns and intensified calls for institutions of higher learning to provide greater access, to demonstrate transparency and accountability, and to be innovative. While the focus has been primarily on undergraduate education, the need to ensure efficient use of the resources devoted to graduate educational programs is equally pressing. The tools of academic analytics [3–5] provide methods for university faculty and administrators in extracting meaning from data regarding a student’s achievements, abilities and aptitude for a particular program [3–5]. The objective of this paper is to document the results of an investigation of the quantitative predictive metrics used in the admissions process by a graduate mechanical engineering program. The hypotheses that motivated this study are:

1. Can an applicant’s undergraduate grade point average (UGPA) and scores on the Graduate Records Examinations (GRE-V, GRE-Q and GRE-W) be used to predict the performance of the applicant in a graduate mechanical engineering program? 2. Is a single construct based on these quantitative predictive metrics a valuable tool in efficiently making unbiased and equitable admissions decisions?

2. Background

The demand for institutions of higher learning to be accountable to all stakeholders has led to investigation of how to assess learning responsibly [1, 6–7]. Stemler [7] addresses the questions of how schools should decide which students to admit and what role testing should play in an admissions process. He
argues that admissions decisions are currently based too heavily on measures of achievement and that that psychometrically sound tools for measuring an applicant’s abilities and aptitude would provide more reliable and robust predictions of performance. The following paragraphs briefly describe each of these characteristics.

Achievement is generally associated with domain-general or domain-specific learning that is gained through formal instruction [8]. The achievement of applicants may be measured by their credentials or through a certification process. Note that achievement is associated with the past and can be documented using historical data such as that provided on academic transcripts.

Ability is a measure of the knowledge and skills currently possessed by an applicant. Ability may be developed through formal training, through apprenticeships or through other life experiences [8–10]. Note that ability is associated with the present, and accurate measurement of ability is independent of the measurement of achievement.

For example, it is expected that a successful applicant to a mechanical engineering graduate program is able to solve differential equations. Consider a case in which an applicant’s academic transcript indicates he or she received an above average grade in a differential equations course, but the applicant is not currently able to solve differential equations. In this case, accurate measurement of the applicant’s achievement would be relatively high while an accurate measurement of the applicant’s ability would be relatively low.

Aptitude is a measure of the applicant’s potential to develop new skills and efficiently acquire new knowledge as circumstances require [7]. Note that aptitude is associated with the future, so measurement of aptitude is challenging. Accurate measurement of aptitude is independent of current abilities or past achievements. It may be argued that assessment of an applicant’s learning orientation and grit may be interpreted as measurement of an applicant’s aptitude. Learning orientation is the extent to which an applicant is open to new ideas in general and is prepared to learn domain-specific skills [7, 9–10]. Grit is the extent to which an applicant has perseverance and passion for long-term goals [11].

The use of high school records and SAT scores for predicting success in college has been extensively studied. Burton and Ramist [12] reviewed many of these studies and found that SAT scores and high school records substantially contribute to prediction of college grades, honors and acceptance by graduate and professional schools. These metrics moderately contribute to prediction of persistence and graduation, and make small but significant contributions to prediction of leadership in college and other accomplishments and to prediction of post-college income.

Similarly, the literature contains many studies or meta-analyses of studies regarding the use of UGPA and standardized test scores as predictive metrics of student performance in a variety of graduate programs. Representative publications are summarized in the following paragraphs.

Kuncel et al. [13] reported the results of a comprehensive meta-analysis of the use of UGPA and GRE scores for graduate student selection and in predicting subsequent performance. Samples representing broad disciplinary areas—humanities, social sciences, life sciences and math-physical sciences—were considered, and academic data for a total of 82,659 graduate students were used. The performance assessments were based on graduate grade point average, first-year graduate grade point average, comprehensive examination scores, publication citation counts, and faculty ratings. The results reported in this study indicated that UGPA and GRE scores are reliable quantitative predictive metrics for the selected performance metrics.

Kuncel et al. [14] conducted another meta-analysis evaluating the potential differential prediction of the GREs for masters versus doctoral programs, and found essentially no differences in the predictive validity of the GRE by program type. In a study commissioned by the Graduate Record Examinations Board, Burton and Wang [15] conducted a collaborative study involving 21 academic departments representing 5 different disciplines from 7 different institutions, resulting in analyzable data for 1,700 students. The authors of this report indicated that attempts to enlist the participation of engineering departments at the 7 institutions were unsuccessful. They attributed the unwillingness of engineering departments to participate to the fact that, at the time data was collected, the demand for engineering graduate students exceeded the supply. Therefore, engineering departments were focused on recruiting enough graduate students to meet their needs rather than on how to select the most qualified applicants. This study was designed to assess the validity of using scores on the quantitative and verbal sections of the GRE and UGPA to predict long-term success in graduate school, where success was measured by cumulative graduate grade point average and faculty ratings of student performance. The disciplines represented in this study were English, Chemistry, Psychology, Education and Biology. The authors of this study also concluded that GRE scores and UGPA were reliable quantitative predictive metrics.

The literature also describes analyses of the usefulness of UGPA and scores from standardized tests typically used in the admissions process at
business, law and medical schools. For example, Kuncel et al. [16] conducted a meta-analysis of the validity of using the Graduate Management Admissions Test (GMAT) scores as a criterion for admission to business school. The data used in this study was obtained from 9 dissertations, 25 journal articles and 12 technical reports, resulting in data for a total of 64,583 applicants. The results of this analysis indicated that the combination of GMAT scores and UGPA are valid metrics for predicting performance in business school. Julian [17] found that MCAT scores significantly increased prediction of medical school grades compared to use of UGPA alone.

Compared to the number of studies regarding the admissions processes employed by professional schools, there are few investigations of the admissions processes used by graduate engineering programs described in the literature. One study described how a two-step approach based on data envelopment analysis (DEA) models was used to assess the appropriateness of selection criteria used by the engineering graduate programs at a single university [18]. The first step included an output oriented DEA model that evaluated and ranked accepted applicants on parameters such as GRE scores, GPA, below-B grades on BS transcripts, and other variables. A second ranking algorithm was implemented to determine student success in the program. The two rankings are then compared to determine the appropriateness of the selection criteria. However, due to concerns regarding the quality of the data used, no definitive conclusions were drawn.

Although the results presented in the studies cited above indicate that UGPA and standardized test results are valid predictive metrics, a large body of literature questions their use or urges that they be used cautiously. Concern over the use of UGPA centers largely on evidence of grade inflation and on the difficulty of comparing grades from institutions with varying standards and policies [19–20]. Concern over the use of standardized test scores is based on questions regarding the predictive ability of these scores [21] and on racial, gender and geographic disparities in these scores [22–25].

Sternberg and Williams [21] considered the validity of the GRE as a predictor of success in a graduate psychology program. Measures of success included grades in the first and second years, faculty evaluation of dissertations and faculty ratings of a student’s creativity and of their research and teaching abilities. The results of this narrow study indicated that GRE scores were useful in predicting grades, but they were of little or no value in predicting the other performance metrics considered.

There is substantial controversy regarding the fair and appropriate use of standardized test scores, and debate of this issue is highly politicized and polemic. Defenders of standardized testing cite a plethora of studies and meta-analyses that indicate standardized tests are unbiased, equitable and valid quantitative predictive metrics and stridently assert that methodologies employed in studies that lead to different conclusions are flawed [26–27]. Critics cite racial, gender and geographic disparities in standardized test scores and assert that standardized tests are ‘crooked measuring sticks’ used by the ‘elites of a mythical meritocracy’ to the disadvantage of various groups [22–25].

Given the controversy regarding the predictive value of an applicant’s UGPA and GRE scores and a lack of studies specifically related to engineering programs, an investigation of the relationship between these quantitative predictive metrics and quantitative performance metrics of those earning an MS from a mechanical engineering program was performed. The analysis documented in this paper resulted in the development of a relationship between the normalized UGPA and GRE scores that is referred to as the Quantitative Application Rating (QAR). The validity and usefulness of the QAR in the admissions process and in the management of other aspects of the graduate program is discussed. Since UGPA and GRE scores are commonly used by engineering departments in the U.S. to make admissions decisions, it is anticipated that the QAR will be valuable to the broader community. The description of the development of the QAR will assist admissions committees in other departments to formulate a similar relationship between quantitative predictive metrics and quantitative performance metrics appropriate to the needs of their department, a process consistent with recommendations for the fair and appropriate use of GRE scores [28].

3. Data collection

3.1 Sample

The sample for this study includes 92 students graduating with MS degrees in mechanical engineering from a private university in the U.S. between April 2010 and August 2012. One student was excluded because of incomplete data. Four non-thesis MS students graduated in that same timeframe and are not included in the study. Eleven students were part of a joint MS/MBA program. Approximately 95% of the students received funding support, with 5% supported exclusively from teaching assistantships, 30% supported by combined teaching assistantships and research assistantships, and 60% supported exclusively by research assistantships and fellowships. Student
and faculty advisor names were replaced by numerical identifiers.

This study is an observational study, meaning that the input variables are not controlled in the study. The values of the input variables are determined by the characteristics of the applicants. Therefore, no individual data point can be replicated. However, the experiment as a whole can be replicated as data is collected from new applicant pools and from students who complete the program.

The predictive metrics (input variables), were extracted from applications submitted to the graduate program. The performance metrics (output variables) were obtained at the MS thesis defense. Each of these variables is described below.

3.2 Quantitative predictive metrics (model input variables)

The input variables are typically used by engineering departments to make graduate admission decisions. Qualitative inputs, such as letters of recommendation and letters of intent, are also commonly used to make admission decisions, but these non-quantifiable inputs are not considered in this study. It is important to note that the set of quantitative predictive metrics is not likely to be comprehensive, so these inputs will not account for all the variance in the quantitative performance metrics. All the input variables are normalized, as will be discussed later. The input variables are summarized in Table 1 and described below:

**UGPA**—Undergraduate Grade Point Average. The grade point averages for this study were based on grades for the last 60 credit hours of university credit on a 4.0 scale. The last 60 credit hours are expected to include a high concentration of courses closely related to the discipline and the most indicative of whether an applicant will be successful in the program. The last 60 credit hours are used throughout the university where the study was conducted, and using this variable can facilitate future possible extension to other disciplines. An alternative approach would be to include all the undergraduate coursework or to include the early science and mathematics courses. The variable $U$ represents the normalized UGPA.

**GREV**—Graduate Records Exam (GRE) Verbal Reasoning Percentile. The GRE is a standardized exam administered by the Educational Testing Service (ETS) and it is commonly used as an admission requirement for graduate programs at U.S. universities, but it is not required for admission to graduate programs at various international engineering programs. The verbal reasoning section is designed to measure an applicant’s “ability to analyze and evaluate written material and synthesize information obtained from it, analyze relationships among component parts of sentences and recognize relationships among words and concepts” [29]. Percentiles were used in this study because the raw score scales are not consistent over the time period of the study. The variable $V$ represents the normalized score on the verbal section of the GRE.

**GREQ**—Graduate Records Exam (GRE) Quantitative Reasoning Percentile. The GRE quantitative reasoning score is designed to measure “problem-solving ability, focusing on basic concepts of arithmetic, algebra, geometry and data analysis” [29]. Again, percentiles rather than raw scores were used. The variable $Q$ represents the normalized score on the quantitative section of the GRE.

**GREW**—Graduate Records Exam (GRE) Analytical Writing Percentile. The GRE analytical writing section is designed to measure “critical thinking and analytical writing skills, specifically your ability to articulate and support complex ideas clearly and effectively” [29]. Percentiles rather than raw scores were also used for this variable. While the Verbal Reasoning and Quantitative Reasoning raw scores vary from 130–170 (after August 2011), the Analytical Writing score range is 0–6, resulting in coarser increments in reported percentiles. The variable $W$ represents the normalized score on the written section of the GRE.

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Normalized Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>UGPA</td>
<td>$U$</td>
<td>Undergraduate GPA</td>
</tr>
<tr>
<td>Input</td>
<td>GREV</td>
<td>$V$</td>
<td>GRE verbal percentile</td>
</tr>
<tr>
<td>Input</td>
<td>GREQ</td>
<td>$Q$</td>
<td>GRE quantitative percentile</td>
</tr>
<tr>
<td>Input</td>
<td>GREW</td>
<td>$W$</td>
<td>GRE writing percentile</td>
</tr>
<tr>
<td>Pooled input</td>
<td>QAR</td>
<td>–</td>
<td>$a_U U + a_V V + a_W W$</td>
</tr>
<tr>
<td>Output</td>
<td>GGPA</td>
<td>$G$</td>
<td>Graduate GPA</td>
</tr>
<tr>
<td>Output</td>
<td>MTG</td>
<td>$M$</td>
<td>Months to graduation</td>
</tr>
<tr>
<td>Output</td>
<td>PR</td>
<td>$P$</td>
<td>Peer-reviewed publication rating</td>
</tr>
<tr>
<td>Output</td>
<td>APR</td>
<td>$A$</td>
<td>Advisor-determined performance rating</td>
</tr>
<tr>
<td>Pooled output</td>
<td>WPR</td>
<td>–</td>
<td>$a_G G + a_M M + a_P P + a_A A$</td>
</tr>
</tbody>
</table>
### 3.3 Performance metrics (model output variables)

The performance metrics were evaluated at the time of the students’ thesis defenses. Each of these output variables is described below.

**GGPA**—Graduate Grade Point Average. The program requires a total of 30 semester hours of credit, of which 24 semester hours are coursework and 6 hours are thesis credit. No common courses are required in the MS program. The thesis credit was graded on a pass-fail scale, so the graduate grade point average is the average grade earned in the 24 semester hours of coursework. The variable $G$ represents the normalized score.

**MTG**—Months to Graduation. The months to graduation represents the number of months from the beginning of the graduate program to the official graduation date. Because some students complete all of their requirements before the official graduation date, this represents an upper bound of the time required to complete the degree. As a performance metric, a smaller time to graduation is preferable to a larger time to graduation, so the variable $M$ is the negative of the normalized score for months to graduation. This metric can be a more ambiguous measure because equally effective graduate students could differ in the length of time they take to graduate for reasons unrelated to graduate school performance (e.g., time off to look after a sick parent, spouse, or child). This and similar considerations may also affect how a program chooses to weight this factor in admissions decisions.

**PR**—Peer-Review Publication Rating. The peer review publication rating is a measure of a student’s productivity in producing peer-reviewed publications of their work by the time of the thesis defense. This rating was used because it is central to the program’s goals of providing valuable mentoring experiences with graduate students, ensures the quality of the work through peer-review, and is consistent with university standards and disciplinary norms for advancement. Students often complete and publish papers after the thesis defense, so plans for peer-reviewed publications are also included in calculation of the peer-review publication rating. The thesis advisor records the number of journal articles, peer-reviewed conference papers, and other conference papers that have been accepted for publication, in review, or are planned for publication at the time of the defense. A peer-review publication rating is calculated by multiplying each type of publication by its corresponding weight, and then summing these values. The weights used for each type of publication are listed in Table 2. The weights are roughly based on the program’s weighting of publications in faculty evaluation of performance. The variable $P$ represents the normalized score.

The mean and standard deviation of the predictive and performance metrics for the 92 students in this study are shown in Table 3.

**APR**—Advisor-determined Performance Rating. At the time of the thesis defense, the thesis advisor rates the student on a scale of 1–5 on 6 different criteria: mastery of topics related to mechanical engineering, understanding of governing principles related to the sub-discipline, ability to conduct independent research, technical writing ability, oral presentation skills, and ethical and moral standards. The average of these individual scores represents the “advisor rating”. The normalized score based on the advisor rating is the Advisor-determined Performance Rating, which is represented by $A$.

**WPR**—Weighted Performance Rating. Each output variable represents a measure of different performance criteria. These can also be combined to create a single Weighted Performance Rating, or $WPR$. The $WPR$ is determined by multiplying each output variable by a weighting factor, then summing. This is shown below, where the weights are represented by $a_i$:

$$WPR = a_G G + a_M M + a_P P + a_A A$$  \hspace{1cm} (1)

### 3.4 Normalization

The statistical models developed for each output are linear functions of the inputs. To make effective models of this type, the data should have the same general shape, although no specific distribution is
needed. In particular, no assumption of normality is placed on either the inputs or the outputs.

To enable direct comparison of the models’ linear coefficients, it is desirable to normalize the data before performing a regression analysis. As all of these distributions have a well-defined central tendency, it is appropriate to normalize them by subtracting the mean and then dividing by the standard deviation, leaving the normalized study variables, whose distributions are shown in Fig. 1:

\[ U_i = \frac{UGPA_i - \bar{UGPA}}{\sigma_{UGPA}} \]  
\[ Q_i = \frac{GREQ_i - GREQ}{\sigma_{GREQ}} \]  
\[ V_i = \frac{GREV_i - GREV}{\sigma_{GREV}} \]  
\[ W_i = \frac{GREW_i - GREW}{\sigma_{GREW}} \]  

Undergraduate GPA (UGPA) shows a central tendency and a relatively symmetric distribution, although there is a slight skew to the left. GRE Quantitative Percentile (GREQ) and GRE Verbal Percentile (GREV) both show a central tendency but skewed to the left. GRE Writing Percentile (GREW) appears to have a bimodal, but symmetric distribution.

The distributions of the performance variables were also investigated. MTG has a central tendency but skew right. GGPA has a central tendency but skew left. APR has a central tendency but skews left. Because these three measures have a central tendency and a moderate amount of skew, it is appropriate to scale them as the predictive metrics were scaled:

\[ M_i = \frac{(MTG_i - \bar{MTG})}{\sigma_{MTG}} \]  
\[ G_i = \frac{GGPA_i - \bar{GGPA}}{\sigma_{GGPA}} \]  

Fig. 1. Distributions of the normalized predictor variables.
As described earlier, $M$ has a negative sign because a large value of $MTG$ is undesirable.

In contrast to the first three metrics, the publication rating shows no central tendency, but appears as an exponential distribution. Therefore, the data were transformed by adding 0.1 and taking the natural logarithm of the result.

$$\log PR_i = \ln(PR_i + 0.1)$$

The normalized peer-reviewed publication rating, $P_i$, is defined as

$$P_i = \frac{\log PR_i - \overline{\log PR}}{\sigma_{\log PR}}$$

Figure 2 shows the distributions of the normalized performance metrics. The normalized data all show a central tendency and are approximately the same in magnitude. These characteristics should lead to reasonable performance when fitting routines are applied.

**Table 4.** Mean, standard deviation, and correlation coefficients for the normalized data. As expected, the means are all zero and the standard deviations are all 1

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Pearson Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
</tr>
<tr>
<td>U: Undergraduate GPA</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q: GRE Quantitative %ile</td>
<td>0</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>V: GRE Verbal %ile</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>W: GRE Writing %ile</td>
<td>0</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>G: Graduate GPA</td>
<td>0</td>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td>M: Months to graduation</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>A: Advisor rating</td>
<td>0</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>P: Publication rating</td>
<td>0</td>
<td>1</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Note: $n = 92$. Correlations greater than ±0.21 are significant at $p < 0.05$ (2-tailed) and correlations ±0.27 are significant at $p < 0.01$. 

Fig. 2. Distributions of the normalized performance variables.
4. Analysis

A statistical analysis exploring the fit of linear models for each of the performance measures was performed. The objective was to find the linear model that best fits the performance metrics to the predictive metrics without spurious over-fitting. Therefore, all possible linear models of the combinations of 1, 2, 3, or 4 of the performance metrics were considered. The analysis was performed using the `regsubsets` function of the `leaps` package [30] in R [31].

An all-subsets exploration of linear regression models for each of the performance metrics was performed. All possible linear models based on 1, 2, 3 or 4 predictive metrics were evaluated. Figure 3 shows the adjusted $R^2$ values for the different models. The “number of predictors” represents the number of input variables used in the linear model. For example, a model that used one variable, such as $U$, is plotted with the symbol “U” with a subset of 1 on the x-axis and the model’s adjusted $R^2$ value on the y-axis. A model using two input variables, such as $U$ and $W$, is shown on the plot as “U–W”, and similarly for three and four variables. For one and two variable models, only the three models with the highest adjusted $R^2$ value for each number of predictive metrics are shown in Fig. 3.

The all-subsets regression analysis shows that the best model for $M$ is a two-term model using $U$ and $W$; the best model for $G$ is a three-term model using $U$, $Q$, and $W$; the best model for $A$ is a two-term model using $Q$ and $W$; and the best model for $P$ is a one-term model using only $Q$. The use of $V$ is not recommended in any of the models, although the second-best model for $G$ includes all four predictive metrics.

Note that it is possible to calculate the quality of the fit using more predictive metrics if desired. For example, the $U-Q-W$ model for $M$ has virtually the same adjusted $R^2$ as the $U-W$ model. However, best practice requires the use of the minimum number of terms.

Fig. 3. Best-fit model factors for each of the performance measures.
The best fit model for \( M \) is a two-term model \( M = U + 2W \). The analysis of variance (ANOVA) results for the regression are shown in Table 5. Since the \( p \)-values of the coefficients for \( U \) and \( W \) are less than 0.05 by nearly an order of magnitude, both terms are significant. Although the \( R^2 \) value for the model is low (14%), the significance of the regression is high (\( p \)-value is 0.001). The values of \( U \) and \( W \) are 0.2777 and –0.2706; the resulting model is:

\[
M = 0.28 U - 0.27 W
\]

(11)

Note that the \( R^2 \) for the model describes the fraction of the variance in the output data (in this case, months to graduation) that is explained by the model. This means that roughly 15% of the variance in time to graduation can be predicted using \( U \) and \( W \). As expected, a higher \( U \) correlates with a shorter time to graduation. It is somewhat surprising that a higher \( W \) score correlates with a longer time to graduation. However, other studies have shown that negative correlations with some GRE scores are not uncommon, in which case it is recommended that the variable be removed from the predictive model [32].

Figure 4 shows the model fit and residual analysis for this model. As we will use these plots for all of our models, the individual plots will be described briefly here. More information can be found in chapter 5 of [33]. Five plots are shown for each model: the model fit, the residuals versus fitted values, the Normal Q-Q of the residuals, the residuals vs. leverage plot, and the residual vs. leverage plot.

The model fit plot shows the actual value of the output \( (M, \text{in this case}) \) as a function of the predicted value for the output. If the model were perfect, all of the data points would lie on the model trend line. The distance between a data point and the trend line is residual for that data point.

The residuals vs. fitted values plot show the model lack of fit as a function of the fitted value. Ideally, there would be no trend in the residuals, either in the mean value or the variation, as the fitted value changes.

The Normal Q-Q plot shows the residual values versus the theoretical value if the residuals were normally distributed. An ideal result would be that all the data lies on the trend line, with the data concentrated around 0.

The Scale-Location plot explores the variance of the residuals as a function of the fitted value. Ideally, there is neither a trend nor a change in the scatter with the fitted value.

The Residuals vs. Leverage plot is used to look for outliers that strongly affect the regression. Leverage describes the distance of a point from the mean of the data, and the standardized residual describes the lack-of-fit for a given data point. Points with high leverage and high residuals may strongly influence the model regression. A measure of the concern about a particular data point is the Cook’s distance. A Cook’s distance of 1.0 indicates a point that may be strongly influential on the regression.

In this particular model, there is no trend in the residuals, the residuals are normally distributed, the scale-location plot is fine, and the highest Cook’s distance is less than 0.5, indicating that the regression is sound.

4.2 Graduate GPA

The best model for \( G \) is a linear model with \( U, Q, \) and \( W (G = \alpha_1 U + \alpha_2 W + \alpha_3 Q) \). The ANOVA results for this model are shown in Table 4. All regression terms are significant. The \( R^2 \) for the model is moderate (42%), but the significance of the regression is high (\( p \)-value is \( 2.2 \times 10^{-10} \)).

The model obtained from the fit is:

\[
A = 0.478 U + 0.236 Q + 0.273 W
\]

(12)
Fig. 4. Model fit and residual analysis for fitted model of M.

Table 6. ANOVA results for linear model of G as a function of U, Q, and W

<table>
<thead>
<tr>
<th>Residuals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>-1.98539</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>U</td>
</tr>
<tr>
<td>Q</td>
</tr>
<tr>
<td>W</td>
</tr>
</tbody>
</table>

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.
Residual standard error: 0.7756 on 88 degrees of freedom.
Multiple R-squared: 0.4182, Adjusted R-squared: 0.3984.
The model fit and analysis of residuals are shown in Fig. 5. The fact that there is less scatter on the right hand side of the plot indicates that the variability of the residuals changes with fitted values but these indications are weak due to a limited number of extreme points. Note that there are very few fitted values above 0.75. The residuals are approximately normally distributed. Four potential outliers are identified, but all have small Cook’s distance values. Therefore, the regression analysis is sound.

### 4.3 Advisor performance rating

The all-subsets regression indicates that the best model for $A$ is a two-factor model containing $Q$ and $W$. The ANOVA results for this model are shown in Table 5. Again, the regression has a low $R^2$ value of 7%, but the regression is still significant at the 95% confidence level, with a $p$-value of 0.032.

The model obtained from the fit is:

$$ A = 0.204 \ Q + 0.154 \ W $$

(13)

The residual analysis is shown in Fig. 6. There are a few potential outliers, but they have minimal effects on the fit, and it can be concluded that this fit is well-behaved.

### 4.4 Publication rating

The all-subsets regression indicates that the best model for $P$ is a one-factor model containing only $Q$. The regression results for this model are shown in Table 6, and the residual analysis is shown in Figure 6. The regression has a low $R^2$ value of 5%, but the regression is still significant, with a $p$-value of 0.031.

The model obtained from the fit is:

$$ A = 0.225 \ Q $$

(14)
Table 7. ANOVA results for linear model of A as a function of Q and W

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.9082</td>
<td>-0.5291</td>
<td>0.1006</td>
<td>0.7047</td>
<td>1.6693</td>
</tr>
</tbody>
</table>

Coefficients:

|       | Estimate  | Std. Error | t value | Pr(>|t|) |
|-------|-----------|------------|---------|---------|
| (Intercept) | 6.788e-16 | 1.014e-01 | 0.000   | 1.0000  |
| Q       | 2.040e-01 | 1.030e-01 | 1.980   | 0.0508  |
| W       | 1.543e-01 | 1.030e-01 | 1.498   | 0.1378  |

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 1.
Residual standard error: 0.9728 on 89 degrees of freedom.
Multiple R-squared: 0.07438, Adjusted R-squared: 0.05358.
F-statistic: 3.576 on 2 and 89 DF, p-value: 0.03208.

Fig. 6. Residual analysis for two-factor fit of A.
There are significant outliers in this model at the low end of the $P$ distribution. This can be seen in the departure of the Normal Q–Q plot from the dashed line. It can also be seen in the three points with a $P$ of -3, but a fitted value of -0.28, 0.7, and 0.25. These results indicate that a few students have extraordinarily low publication ratings compared with what would be expected based on their score on the
quantitative section of the GRE. For example, data point 14 has a relatively high Cook’s distance score. However, this value is still below the concern threshold of 1. Therefore, it is concluded that the regression is reliable.

Although it is not critical for making the admissions decision, in terms of achieving graduate success it may be profitable to carefully examine the students on the low end of the P distribution, as they appear to have a different distribution than the rest of the students. The low-P points appear to have a different slope on the Normal Q–Q plot than the rest of the points.

4.5 Weighted performance rating (WPR)

The WPR is a pooled measure of success. We define WPR with $a_G = a_M = a_P = 2$ and $a_A = 1$, because we choose to place more weight on the objective performance metrics than on the subjective APR. Many different rationally chosen weights are possible. The weights used may be varied to reflect a combination of the confidence in the data and the program’s values. The histogram for WPR is shown in Fig. 8. There is a central tendency, but the distribution is skewed to the left, with some relatively large negative values. This likely indicates that some students are evaluated poorly relative to multiple performance metrics.

The all-subset regression identifies a two-term model for WPR containing only $U$ and $Q$ as the optimum model, as shown in Fig. 9.

The results for the regression of this model are shown in Table 7 and the residual analysis for the regression is shown in Fig. 10. While there are a few potential outliers, the Cook’s distance metric indicates that these points are unlikely to negatively affect the fit.

The model obtained by this fit is:

$$WPR = 1.7 \ U + 1.3 \ Q$$

(15)

Note that neither $W$ nor $V$ is significant in fitting this model.

5. Discussion

Table 3 summarizes the models resulting from the regression analysis. In addition to the formal predictive model, we also choose a scaled predictive model. The scaled predictive model uses integer coefficients, with the objective of making it easy for those seeing the model to understand the relative weights. Because we are only interested in evaluat-
ing the predicted success of candidates relative to one another, the absolute magnitude of the coefficients is unimportant. We therefore choose integer coefficients whose relative magnitudes are close to those in the predictive model.

It is interesting to note that $V$ occurs in none of the models. $U$ is in two of the individual measures and in the combined measure. $Q$ occurs in three of the individual measures and the combined measure. $W$ is in three of the individual measures (once with a negative coefficient), but it does not occur in the combined measure.

It may seem surprising that $W$ is not significant in the combined measure, even though it is significant in three of the individual measures. This is due to the fact that $W$ is the only factor that has a negative coefficient in any of the models. The negative coefficient in $M$ offsets the positive coefficient in $G$, and reduces the overall contribution of $W$ to $WPR$.

Prior to the study, intuition predicted that $W$ would be an indicator of the publication rating. The counter-intuitive result suggests that the publication rating may be more based on the students' ability to generate the research content for the

<table>
<thead>
<tr>
<th>Residuals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
</tr>
<tr>
<td>–17.0309</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>$U$</td>
</tr>
<tr>
<td>$Q$</td>
</tr>
</tbody>
</table>

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.
Residual standard error: 4.1 on 89 degrees of freedom.
Multiple R-squared: 0.2333, Adjusted R-squared: 0.2161.
F-statistic: 13.54 on 2 and 89 DF, p-value: 7.351e-06.
paper than on skills measured by the GRE writing exam.

5.1 Calculated index (quantitative application rating—QAR)

One objective of this work was to evaluate whether a single construct based on the quantitative predictive metrics could be a valuable tool in efficiently making admissions decisions. The scale models could be used as one element of the evaluation of graduate school applicants. The scaled model can also be used as a basis for other models that may be employed for comparing applicants. An example of such a model is provided here and is in use at the authors’ institution, and this can serve as a guide for other institutions for developing their own models.

In this approach, a Quantitative Application Rating (QAR) is calculated for each applicant. The QAR combines all the inputs into one model.

![Graphs showing residual analysis for two-factor model of WPR.](image)

**Fig. 10.** Residual analysis for two-factor model of WPR.

**Table 10. Summary of regression models**

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Predictive Model</th>
<th>Suggested Scaled Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Months to Graduation</td>
<td>( M = 0.28U - 0.27W )</td>
<td>( M = U - W )</td>
</tr>
<tr>
<td>Graduate GPA</td>
<td>( G = 0.48U + 0.24Q + 0.27W )</td>
<td>( G = 2U + Q + W )</td>
</tr>
<tr>
<td>Publication Rating</td>
<td>( P = 0.23Q )</td>
<td>( P = Q )</td>
</tr>
<tr>
<td>Advisor Performance Rating</td>
<td>( A = 0.20Q + 0.15W )</td>
<td>( A = Q + W )</td>
</tr>
<tr>
<td>WPR (2G+2M+2P+A)</td>
<td>( WPR = 1.7U + 1.3Q )</td>
<td>( WPR = 5U + 4Q )</td>
</tr>
</tbody>
</table>
and GRE Verbal scores are not high, some practical factors suggest that they should be included in application decisions. The first is to avoid inadvertently implying that writing and verbal ability are unimportant for success in a mechanical engineering graduate program. Second, the university’s internal comparison of graduate programs includes evaluations of all GRE scores. Therefore, we have elected to create a model that combines the predictive model with minimally weighted values of the GRE Verbal and GRE Written scores. The following equation results:

$$QAR = 5U + 4Q + V + W$$  

or

$$QAR_i = 5\left(\frac{UGPA_i - \bar{UGPA}}{\sigma_{UGPA}}\right)$$  

$$+ 4\left(\frac{GREQ_i - \bar{GREQ}}{\sigma_{GREQ}}\right)$$  

$$+ \left(\frac{GREVi_i - \bar{GREV}}{\sigma_{GREV}}\right)$$  

$$+ \left(\frac{GREWi_i - \bar{GREW}}{\sigma_{GREW}}\right)$$  

In the evaluation of graduate applicants, a QAR value is calculated for each applicant. The averages (UGPA, GREQ, GREV, GREW) and standard deviations ($\sigma_{UGPA}$, $\sigma_{GREQ}$, $\sigma_{GREV}$, $\sigma_{GREW}$) are determined from the population of all currently enrolled students. Applicant undergraduate GPA and GRE scores are added to a spreadsheet during the application process. These scores are carried onto another spreadsheet when students enroll, making these calculations straightforward.

A QAR value for an applicant provides a comparison relative to the current set of graduate students in the program. A value of $QAR_i = 0$ represents an applicant with predictor variables on par with students currently in the program, while positive scores are above average and negative scores are below average compared to currently enrolled graduate students.

As a check for the validity of this QAR formulation, the four-factor QAR was fit to WPR data for completed students. In addition, a two-factor reduced QAR0 was developed and fit to WPR data:

$$QAR0 = 5U + 4Q$$  

or

$$QAR0_i = 5\left(\frac{UGPA_i - \bar{UGPA}}{\sigma_{UGPA}}\right)$$  

$$+ 4\left(\frac{GREQ_i - \bar{GREQ}}{\sigma_{GREQ}}\right)$$  

Both regressions are highly significant ($p \sim 2 \times 10^{-6}$). Both have relatively low $R^2$ values (about 22%). The two-factor QAR0 has higher leverage data points at a lower Cook’s distance than does the four-factor QAR. However, none of the differences between these models appear to have practical significance. Therefore, continued use of the four-factor model to communicate the importance of written and verbal communication to the students is justified.

5.2 Limitations

While this work reflects the best analysis that could be performed on the existing data, it is nonetheless limited by the available data. The major limitations observed by the authors are discussed here.

Reliability. No study has been performed on the reliability of the Advisor Performance Rating. As it is a subjective rating, and performed by different individuals for different students, there is a possibility of rater bias. Coded data on the raters is part of the data set, and future plans call for an analysis of

Table 11. ANOVA results for linear model of WPR as a function of QAR

<table>
<thead>
<tr>
<th>Residuals:</th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>–17.5557</td>
<td>1.6647</td>
<td>0.6273</td>
<td>2.8379</td>
<td>9.6050</td>
<td></td>
</tr>
</tbody>
</table>

| Coefficients: | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------|----------|------------|---------|---------|
| (Intercept)   | –3.055e-16| 4.280e-01 | 0.000   | 1       |
| QAR           | 3.003e-01 | 5.913e-02 | 5.078   | 2.04e-06***|

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 1.
Residual standard error: 4.106 on 90 degrees of freedom.
Multiple R-squared: 0.2227, Adjusted R-squared: 0.2141.
F-statistic: 25.79 on 1 and 90 DF, p-value: 2.044e-06.
this data. However, given the number of different raters (24) compared to the number of students (92), there is not yet enough data to appropriately assess the reliability of the APR. The significant correlations between APR, PR, and GGPA indicate that the APR does assess the quality of the student’s performance.

**Range Restriction.** The students whose performance was evaluated for this study represents students who matriculated in the program and

---

**Table 12. ANOVA results for linear model of WPR as a function of QAR0**

<table>
<thead>
<tr>
<th>Residuals:</th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-17.0583</td>
<td>-1.8411</td>
<td>0.3812</td>
<td>2.8717</td>
<td>8.3663</td>
</tr>
</tbody>
</table>

| Coefficients:       | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------|----------|------------|---------|---------|
| (Intercept)         | -6.031e-16 | 4.251e-01 | 0.000   | 1       |
| QAR0                | 3.355e-01  | 6.412e-02 | 5.232   | 1.09e-06 *** |

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.
Residual standard error: 4.078 on 90 degrees of freedom.
Multiple R-squared: 0.2332, Adjusted R-squared: 0.2247.
F-statistic: 27.37 on 1 and 90 DF, p-value: 1.085e-06.

---

**Fig. 11. Residual analysis for WPR on four-factor QAR.**
graduated over the course of the study. Students who were not accepted to the program are not included. As acceptance was based partially on the predictive measures considered in this study, the predictive measures show less variation in the study population than in the applicant population. Therefore, it is not possible to generalize this study to the larger applicant pool. However, because the study population shows less variation than the applicant population, it is likely that the estimates of the relationships between the predictive and the performance metrics are downwardly-biased. Hunter et al. [34] describe the formulas necessary to calculate the effects of the indirect range restriction on this study. Because the data necessary to perform these corrections is not available, the uncertainty due to range restriction remains.

Sample Size. Although this sample represents all of the M.S. thesis students graduating from the mechanical engineering program during the course of the study, the size of the sample is relatively small. Given this limitation, caution should be used when applying the results of this study to larger populations and other academic departments.

6. Summary
Graduate school admission decisions have profound effects on students, universities, and the quality of the educational experience. Admissions recommendations are generally based on both subjective measures and quantitative metrics of an applicant’s credentials and qualifications. Academic analytics can assist university faculty and administrators in extracting meaning from the quantitative predictive metrics provided by applicants to graduate programs. This paper documents the results of an investigation of the extent to which
an applicant’s UGPA and GRE scores predict the performance of students admitted to a graduate mechanical engineering program and of the usefulness of a single construct based on these quantitative predictive metrics as a tool for making unbiased and equitable admissions decisions.

Regarding the first hypothesis, it was found that some quantitative predictive metrics available at the time of application to an engineering graduate program are valid predictors of graduate school performance. For the data and metrics presented, the dominant predictors of the selected performance metrics are undergraduate GPA and the score on the quantitative section of the GRE. The formula for predicting overall performance in the program studied is \( 5U_i + 4Q_i \), where \( U_i \) and \( Q_i \) are the normalized scores of undergraduate GPA and GRE quantitative percentile based on several years of admitted students. The GRE writing score is correlated with an increased time to graduation and an increase in graduate GPA. The GRE verbal score is not a significant predictor of any of the selected performance metrics.

Regarding the second hypothesis, it was found that the Quantitative Application Rating (QAR) is a valid construct that facilitates admissions-related tasks. These results can serve as a baseline model for adapting the QAR to other programs that may have different gender or racial compositions, that may be more or less competitive in the admissions process, or programs that value different weightings for the performance metrics.

7. Conclusions

It is important to note that the quantitative predictive metrics only account for a fraction of the variance in the performance metrics. These results demonstrate that success in the graduate engineering program studied depends on factors not captured by these quantitative predictive metrics. It is probable that the UGPA accurately characterize an applicant’s achievement and that the GRE quantitative score accurately characterizes an applicant’s ability, but these metrics do not measure an applicant’s aptitude. Quantitative measurement of an applicant’s aptitude is challenging, so use of subjective measures—letters of recommendation, prior research experience, and interview results—to assess an applicant’s aptitude for the program is recommended.

The quantitative measures (particularly the QAR) are currently being used in several ways in our graduate program, some of which are described here. First, when combined with subjective measures, the QAR is a valuable tool in making admissions decisions. Second, it is a primary factor in objectively identifying students that should be awarded department fellowships. Third, average scores for an applicant cohort is compared to averages for applicant pools from previous years, providing an opportunity to study the effectiveness of efforts to recruit well prepared students. Fourth, these metrics are used to compare cohorts of admitted and matriculated students and to evaluate the effectiveness of efforts to recruit the most highly qualified applicants.

Since many graduate engineering programs in the U.S. use UGPA and GRE scores as a guide in the admissions process, it is anticipated that the results presented in this paper will be of value to the broader community. It is relatively straightforward to modify the weights used in the performance metrics described or to incorporate additional performance metrics. Therefore, graduate programs may use the analysis described in this paper as a model for validating the use of predictive metrics or for developing performance metrics that reflect the unique objectives of their program. The results obtained in this study could be a starting point for more general results that extend to other engineering disciplines and other institutions. It is anticipated that the description of the QAR will aid other departments in developing a similar relationship between quantitative predictive metrics and quantitative performance metrics appropriate to the needs of their department, a process which is recommended for the fair and appropriate use of GRE scores [28].

Acknowledgments—The assistance of Miriam Busch, Graduate Advisor for the Department of Mechanical Engineering at Brigham Young University, is gratefully acknowledged.

References

Laryl H. Howell is a professor and past chair of the Department of Mechanical Engineering at Brigham Young University (BYU), where he also holds a University Professorship. Prior to joining BYU in 1994 he was a visiting professor at Purdue University, a finite element analysis consultant for Engineering Methods, and an engineer on the design of the YF-22 (the prototype for the U.S. Air Force F-22). He received his PhD and MS degrees from Purdue University and his BS from Brigham Young University. He is a licensed professional engineer and the recipient of a National Science Foundation CAREER Award, a Theodore von Kármán Fellowship, the BYU Technology Transfer Award, the Maeser Research Award, several best paper awards, and the ASME Mechanisms & Robotics Award. He is a Fellow of ASME, past chair of the ASME Mechanisms & Robotics Committee, past co-chair of the ASME International Design Engineering Technical Conferences, and a past Associate Editor for the Journal of Mechanical Design. Prof. Howell’s technical publications and patents focus on compliant mechanisms, including origami-inspired mechanisms, micro-electromechanical systems, and medical devices. He is the author of the book Compliant Mechanisms published by John Wiley & Sons.

Carl D. Sorensen is Professor of Mechanical Engineering and is currently the holder of the B. Keith Duffing Teaching and Learning Fellowship. He received a B.S. in Physics from Brigham Young University, where he was elected a member of Sigma Pi Sigma, the national physics honor society. He received a Ph.D. in Materials Science from the Massachusetts Institute of Technology, where his dissertation involved digital signal processing in gas-tungsten arc welding. Following postdoctoral work at MIT and with Chrysler and General Electric Aircraft Engines, he began teaching at BYU. During his tenure at BYU, he has served as a faculty member in Manufacturing Engineering as well as Mechanical Engineering. He has also served as a Visiting Professor of Design at the Kanazawa Institute of Technology, in Kanazawa Japan.

Matthew R. Jones is an Associate Professor of Mechanical Engineering at BYU where he teaches courses in the areas of heat transfer and thermodynamics. Currently, he is involved in research projects related to reduced order modeling of turbomachinery and thermal processes; analysis, design and optimization of clean-burning, fuel-efficient, modular biomass cookstoves, power harvesting, optical fiber thermometry and measurement of thermophysical properties. Prior to coming to BYU, he was an Assistant Professor in the Department of Aerospace and Mechanical Engineering at The University of Arizona, and a Science and Technology Agency Fellow at the Mechanical Engineering Laboratory in Tsukuba, Japan. Professor Jones has also held research appointments at the Marshall Space Flight Center (NASA) and at Argonne National Laboratory.