Composite Measures of Academic Performance: Insights from General Systems Performance Theory

George V. Kondraske
Department of Electrical Engineering, Department of Bioengineering
University of Texas at Arlington

Lawrence R. DiSalvi
Department of Bioengineering
University of Texas at Arlington

Abstract

Quantitative characterization of academic performance and competency have been topics of interest at all educational levels. Efforts generally include tests within major subject categories such as math, reading, and science with the diversity of categories increasing at higher educational levels. Such tests result in a score (for individuals or aggregated across individuals) for each category that is intended to reflect "performance"; i.e., a level of competency or mastery within the defined subject matter area (SMA). While individual SMA scores are used as part of individual and group performance reports, single number "overall performance" metrics are frequently also used. Such composite measures have traditionally been based on simple or weighted averaging. General Systems Performance Theory (GSPT) suggested a fundamentally different, more conceptually sound, and intuitively attractive approach, based on the notion of a performance capacity envelope and computation of its volume. This paper presents the application this approach to the issue of composite measure formation in education contexts using realistic, illustrative example cases. It is argued that the quantitative "amount" of academic performance capacity that is the intended characteristic of interest is more accurately reflected with the GSPT-based approach.

Introduction

Mathematical averaging of course grades has been a long-standing, traditional method to achieve a composite score reflecting an individual's or defined population's scholastic performance[1,2,3]. Composites such as grade point averages (GPAs) and the mathematical sum of test result components on standard college entrance examinations serve as a basis for decision-making in job hiring and advanced educational admissions. Calculation of grade point average (GPA) is standard procedure in most schools. One purpose assigned to GPAs is to predict which students will successfully complete a desired objective, such as graduating from college or performing a particular job. Colleges routinely use the student's high school GPA and college entrance examinations in making admissions decisions. Yet, studies over many years have consistently shown that such measures are poor predictors of success[2,3,4,5,6,7].

Any GPA *is* an average and therefore its computation requires the use of addition. However, this traditional computation ignores units of measure. For example, the undefined and therefore unstated (but *nonetheless present*) units of math competency *("Einsteins"*, perhaps) must be added to the unstated and present units of written composition competency *("Hemingways"*?) in the computation of a GPA involving those two SMAs. Other frequently used composites such as SAT and ACT overall scores incorporate the same additive process which implies that math and written composition competencies are interchangeable and a deficiency in one can be overcome by a surplus in the other. This additive process is akin to computing an average with measures of units of physical force (newtons) with units of distance (meters) and disregarding the concept that these represent unique quantities.

Addition-based composite measures of performance are found in many fields of endeavor. To compute an average of different quantities when units are actually recognized or impossible to ignore, normalization is often employed (i.e., dividing a measure by a reference value with the same units to make it appear that the units have "disappeared". However, they remain present *conceptually* (i.e., a math competency still represents math competency which is not the same as writing composition competency).

Interestingly, GPAs have not only been used as predictors of success, but also as outcome measures; e.g., measures of academic success in higher education[8,9]. It is argued that this further conflicts attempts to develop predictive models of success. Many factors have been suggested regarding the relatively poor predictive power of GPA and similar measures of academic performance. Continued manipulation of the variables using the same basic concept is not likely to result in improvements. A fresh perspective is warranted.

Based on more recent developments in performance theory[10,11], the most frequently used methods of composite score formation are argued to be flawed conceptually. An alternate approach resulting in a grade point product (GPP) is presented and discussed. New measures are demonstrated to be conceptually sound and are furthermore suggested to better reflect actual overall competency in a given course of study.

General Systems Performance Theory

Motivated not by a need or desire for aggregate measures, but rather as a basis for understanding human performance, Kondraske introduced GSPT[10,11]. Its stated objectives[11] are to provide: 1) a conceptual basis to define and measure all aspects of any system's performance; 2) a conceptual basis to analyze any task and facilitates system-task interface assessments; and 3) identification of the principles that explain success/failure in any given system-task interface. Briefly, all unique qualities describing a system's capacity to execute its function are modeled using a resource construct. A major GSPT concept is that of a performance capacity envelope (PCE). Each type of performance resource (e.g., speed, accuracy, strength, range, math skill, communicatin skill, etc.) is asserted to represent one dimension of performance (DOP) in a multi-dimensional performance space. Measures of the amount available (R_A) for each performance resource in this space define the systems's performance capacity envelope (PCE).

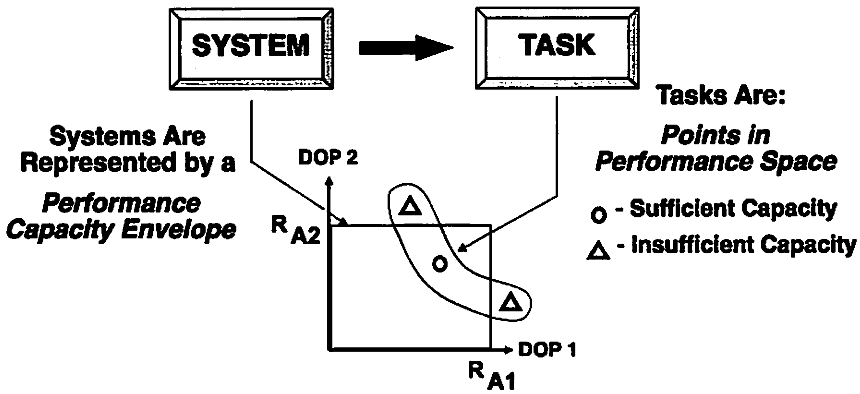

Figure 1. Key GSPT constructs are illustrated for illustrative case considering only two dimensions of performance (DOP) such as math (DOP 1) and communication (DOP 2) competency. The system is characterized by performance resource availabilities (R_{A1}, R_{A2}) that define a performance capacity envelope. Tasks are points in this space, each of which requires a specific amount of each performance resource. The area (or volume) of the envelope represents the "number of points enclosed"; which reflects *the capacity of that system to execute tasks that make demands on the performance resources in this space.*

Once system performance is characterized in this way, the PCE volume (i.e., mathematical product of R_As for a simple representation of any given system) is a rather obvious composite performance measure. Kondraske and colleagues have suggested that this provides a strong conceptual framework for composite performance measure formation and have demonstrated its use in a range of applications including neuromotor control[12,13,14], human motion quality/coordination[15,16,17], disease severity[18], brain performance capacity[19], biomechanics[20,21], diamond quality[11], and education/training[22,23].

Methods

To examine and contrast the behavior of GPP relative to GPA composites, illustrative cases were developed using hypothetical, realistic data. A data set consisting of competency values for five different Subject Matter Areas (SMAs) and for five different representative students was created. Specific SMA types are not relevant. For the present purpose, it is only necessary to assume that they represent different competencies. For illustrative purposes, some well-known areas (i.e., math, reading, writing, science, etc.) are identified. Each of the five example students are defined to have competency scores covering a range from 60 to 100 on a 0-100 scale. A set of n-dimensional GPA and GPP scores are computed, where "n" represents the number of SMAs included.

A separate analysis was conducted to explore the behavior of GPP vs GPA in the context of a "change in competency level". Baseline values of 75 out of 100 are assigned to each of the five individual SMAs. Two levels of relatively small change (i.e., 5% and 10%) relative to baseline levels are considered. Once again, various n-dimensional GPA and GPP scores are computed and then the percent change from the respective composite baseline (i.e., GPA or GPP) is computed.

Results

Data for individual SMA areas and computed n-dimensional GPA and GPA composites are shown in Table 1.

Table 1. Individual Subject Matter Area (SMA) scores and corresponding GPA and GPP composite scores for five students representing different levels of SMA competency.

Measure ID	Student A	Student B	Student C	Student D	Student E
SMA1 (e.g., Math)	100	90	80	70	60
SMA2 (e.g., Reading)	100	90	80	70	60
SMA3 (e.g., Writing)	100	90	80	70	60
SMA4 (e.g., Science)	100	90	80	70	60
SMA5 (e.g., Computing)	100	90	80	70	60
GPA - 2D*	100	90	80	70	60
GPP - 2D**	100	81	64	49	36
GPA - 3D*	100	90	80	70	60
GPP - 3D**	100	73	51	34	22
GPA - 4D*	100	90	80	70	60
GPP - 4D**	100	66	41	24	13
GPA - 5D*	100	90	80	70	60
GPP - 5D**	100	59	33	17	8

*GPA_2D = AVERAGE(SMA1, SMA2), GPA_3D = AVERAGE(SMA1, SMA2, SMA3), etc.

**GPP_2D = $(SMA1*SMA2)/100^2)*100$, GPP_3D = $(SMA1*SMA2*SMA3)/100^3)*100$, etc.

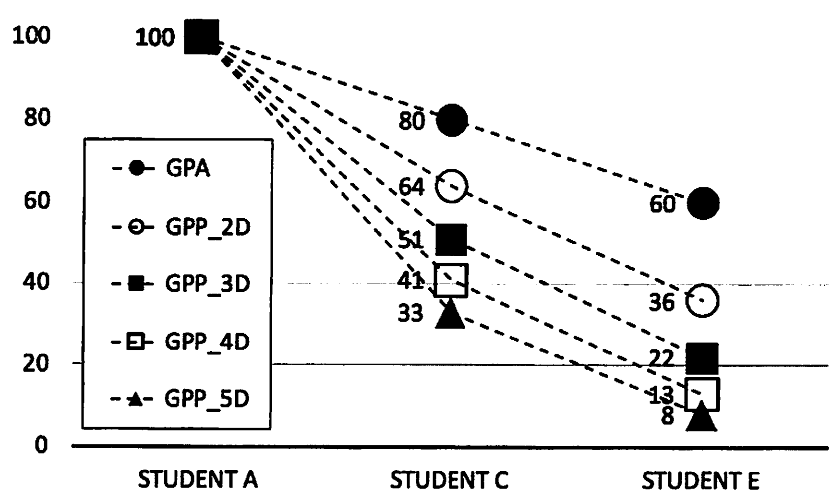

Figure 2. Comparison of the relative levels of competency communicated by GPA (same for 2D thru 5D cases) and GPP composites for three of the five hypothetical student competency levels considered.

Table 2. Baseline and "change from baseline" scores in raw and "percent change" forms for SMAs, computed GPAs (2D, 3D, 4D, and 5D are the same), and computed GPPs.

Measire ID	Raw Measures				Percent Change from Baseline		
	Baseline	Case 1 (5% SMA Baseline Δ)	Case 2 (10% SMA Baseline Δ)		Baseline	Case 1 (5% SMA Baseline Δ)	Case 2 (10% SMA Baseline Δ)
SMAn	75	71.3	67.5		0	-5	-10
GPA - nD	75	71.3	67.5		0	-5	-10
GPP - 2D	56	50.8	45.6		0	-9	-19
GPP - 3D	42	36.2	30.8		0	-14	-27
GPP - 4D	32	25.8	20.8		0	-19	-35
GPP - 5D	24	18.4	14.0		0	-23	-42

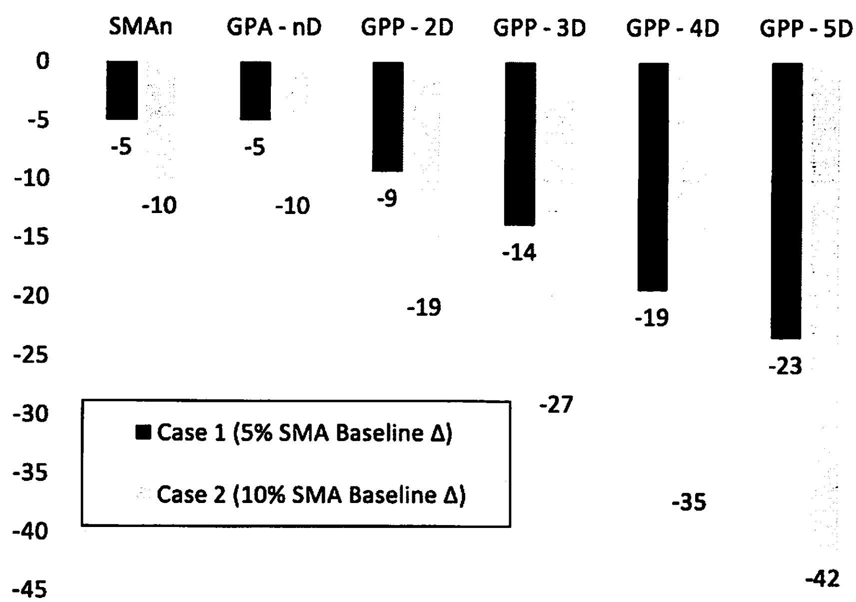

Figure 3. Graphical illustration of "percent change from baseline" scores for: 1) individual SMAs (levels for all five SMAs are the same), 2) GPA – nD (2D, 3D, 4D, and 5D are the same), and 3) GPPs for 2D through 5D cases. Two cases are considered (5% and 10% decreases in each of the five *individual SMA competencies*) giving rise to GPA and GPP scores from which "percent changes from baseline" are then computed.

Discussion

Data in Table 1 and Figure 2 illustrate that GPA and GPP composites communicate very different levels of competency, with GPP measures appearing to be more sensitive. Note that GPAs for the five representative students are same for each competency level regardless of whether two, three, four, or five SMAs are included. Comparing each GPP – nD value to the corresponding GPA – nD value, one finds that the GPP communicates a rather dramatically less level of integrated composite performance capacity. Recall that, as per GSPT (see Figure 1), the volume of the performance capacity envelope is asserted to represent the capacity of the system (i.e., a student) to execute tasks that draw upon the performance resources that define the system in performance space (i.e., math *and* reading, math *and* reading *and* writing, etc.). Which measure better reflects the true level of competency? Does the student with scores of 60 in each of the five SMAs have 60% (GPA – 5D) or 8% (GPP – 5D) of the capacity to do tasks that draw upon all five competencies (such as engineering). To further support the conceptual basis of this performance capacity envelope interpretation, the connection to joint probability (i.e., probability of having sufficient math and reading and writing competency) has been discussed elsewhere[18,23].

It is noteworthy that both GPA and GPP representation, computed with the same set of SMA values will provided the same rank order. But again, rank order of some measure and capacity to execute tasks convey very different types of information.

This begs the question, "What does a GPA *really* represent?", versus what it has been somewhat relentlessly *expected to represent*. Quite simply, it reflects average academic performance or competence across the constituent SMAs. It is "forgiving" (misleading?) in the sense that it allows substitution of one competency for another. In doing so, it essentially disregards the notion of "pre-requisites" and that a certain amount of a competency *in each* of a specific set of SMAs is required for success in complex task situations such as in an engineering educational program or job. Which type of composite measure, GPA or GPP, embraces the idea that a given amount of competency in one SMA is a necessary *but not sufficient* requirement for success in such task contexts?

Table 2 and Figure 3 were included to represent a situation that is often of keen interest in education: How are we doing? Are we getting better or worse and – more significantly in the present context – *by how much*? This type of consideration is nicely reflected by the well-known "Nations Report Card"[24]. When the report indicates a small drop (e.g., 5%) in each of several SMAs, is *the impact* (i.e., the capacity to execute tasks needing those competencies) also only 5% (as would be reflected by GPA) or much greater (as would be reflected by GPP measures) as suggested by Figure 3? While Table 2 and Figure 3 consider declines in group competency levels, the examples also suggest that small *improvements* in SMA competencies can have substantial improvements in complex task performance capacities. The GSPT performance capacity representation indicates improvement in one SMA competency *does not add to*, *but rather multiplies* other competencies, to impact capacity to execute tasks that require multiple SMA competencies.

The desire and need to aggregate measures of a similar character, especially those related to performance or quality, is no different in education than it is in many other fields. Despite the almost universal interest and ubiquitous appearance of addition-based composites, there is rather sparse literature available regarding the generic issue of composite measure formation. One interesting exception is in the field of global economics[25], which describes the use of both the arithmetic mean (additive) and the geometric mean (multiplicative) to form aggregate measures. It may be of interest and insight to readers that both have been used at different points in time to compute the widely publicized United Nations' Human Development Index[26].

In addition to the applications of GSPT's performance capacity envelope concept and product-based composites by Kondraske and colleagues, others have also taken note of and adopted the GSPT perspective on composite formation[27-31].

Conclusions

Conceptual arguments differentiate the appropriateness of the GSPT-based composites over other approaches incorporating mathematical addition. Even if unstated, units of measure for components are recognized and retrained in the final composite so that an intuitive sense of the quantities measured, and dimensionality, is not lost. While units of measure are not used currently for math, reading, science, and other SMA competencies, perhaps they should be. The stated interpretation of n-D GPPs is intuitive and supported by join probability. While scaling of measures used to characterize individual SMA competencies should be consistent across dimensions, no normalizations or transformations are required.

Cited work in other application areas has addressed the issue of validating which type of measure (addition-based or product-based) better reflect true performance capacity. Product-based approaches have shown better correlation with gestalt human observation of performance ein complex tasks. Additional work is recommended to address such validation in educational contexts. It is suggested that many researchers are likely to have existing data sets that can be revisited from the GSPT performance capacity envelope perspective.

The notion of a performance envelope can be traced to aircraft performance characteerization in the 1940s[32]. GSPT generalizes the concept to any and all systems, including human systems. Relative to use of GPAs without justification, it is argued to be more of a *systems engineering approach* to more meaningfully characterizing the aggregate collective capacities the creation and nurturing of which represent the ultimate goals of educational efforts.

References

1. Marbouti, F., Ulas, J. and Wang, C.H., 2020. Academic and demographic cluster analysis of engineering student success. IEEE Transactions on Education, 64(3), pp.261-266.
2. Bretz Jr, R.D., 1989. College grade point average as a predictor of adult success: A meta-analytic review and some additional evidence. Public Personnel Management, 18(1), pp.11-22.
3. Soh, K.C., 2010. Grade point average: what's wrong and what's the alternative? Journal of Higher Education Policy and Management, 33(1), pp.27-36.
4. Imose, R. and Barber, L.K., 2015. Using undergraduate grade point average as a selection tool: A synthesis of the literature. The Psychologist-Manager Journal, 18(1), p.1. online publication. http://dx.doi.org/10.1037/mgr0000025
5. Patacsil, F.F., 2020. Predicting University Students' Academic Success Using Different Tree Classifiers And Ensemble Approaches To Suggest Suitable Program. International Journal of Scientific & Technology Research, 9(2), pp.6001-6009.
6. Blandford, D., Randall, M. and Lofton, J., 2019. Course and Exam Statistics in Electrical Engineering. ASEE IL-IN Section Conference. 5. https://docs.lib.purdue.edu/aseeil-insectionconference/2019/assess/

7. Rodríguez, J.V., Bueno-Delgado, M.V., Sánchez-Aarnoutse, J.C., Pastor-Franco, J.Á. and Juan-Llácer, L., 2019. Impact of admission grade point average on academic results of university students in electrical engineering. The International Journal of Electrical Engineering & Education, p.0020720919874066.
8. Abdelfattah, F.A., Obeidat, O.S., Salahat, Y.A., BinBakr, M.B. and Al Sultan, A.A., 2021. The predictive validity of entrance scores and short-term performance for long-term success in engineering education. Journal of Applied Research in Higher Education. https://www.emerald.com/insight/content/doi/10.1108/JARHE-04-2021-0126/full/html
9. Lackey, L.W., Lackey, W.J., Grady, H.M. and Davis, M.T., 2003. Efficacy of using a single, non-technical variable to predict the academic success of freshmen engineering students. Journal of Engineering Education, *92*(1), pp.41-48.
10. Kondraske, G.V., 1987. March. Human performance: Science or art. In Proceedings of the Thirteenth Northeast Bioengineering Conference (pp. 44-47).
11. Kondraske, G.V., 2011. General systems performance theory and its application to understanding complex system performance. Information knowledge systems management, 10(1-4), pp.235-259.
12. Behbehani, K., Kondraske, G.V. and Richmond, J.R., 1988. Investigations of upper extremity visuomotor control performance measures. IEEE Transactions on Biomedical Engineering, 35(7), pp.518-525.
13. Jafari, M., Wong, K.H., Behbehani, K. and Kondraske, G.V., 1989. Performance characterization of human pitch control system: an acoustic approach. The Journal of the Acoustical Society of America, 85(3), pp.1322-1328.
14. Kondraske, G.V., 1989, November. Neuromuscular performance: Resource economics and product-based composite indices. In Images of the Twenty-First Century. Proceedings of the Annual International Engineering in Medicine and Biology Society, (pp. 1045-1046). IEEE.
15. Fischer, C.A. and Kondraske, G.V., 1997, October. A new approach to human motion quality measurement. In Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136) (Vol. 4, pp. 1701-1704). IEEE.
16. Kondraske, G.V. and Vasta, P.J., 2000, July. Neuromotor channel capacity, coordination, and motion quality. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143) (Vol. 3, pp. 2259-2262). IEEE.
17. DiSalvi, L.R. and Kondraske, G.V., 2021. Speed-Accuracy Trade-Offs and General Systems Performance Theory: Novel Application to Fitts' Law and Beyond. In 37th International Symposium on Aviation Psychology (p. 214).
18. Kondraske, G.V. and Stewart, R.M., 2006, January. Quantitative characterization of disease severity in diseases with complex symptom profiles. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 3966-3969). IEEE.
19. Saganis, X., Goreczny, A. J., Kondraske, G., Berman, M., Comnick, I., Allen, T., & Nussbaum, P. D. (2021). Test–retest reliability of RC21X: a web-based cognitive and neuromotor performance measurement tool. Archives of Clinical Neuropsychology, 36(3), 322-328.
20. Kondraske, G.V., 1987. Computation of functional capacity: Strategy and example for shoulder. In Proc. 9th Annual IEEE Eng. in Med. and Biol. Soc. Conf (pp. 477-478). IEEE.
21. Vasta, P.J. and Kondraske, G.V., 1997, October. An approach to estimating performance capacity envelopes: knee extensor system example. In Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136) (Vol. 4, pp. 1713-1716). IEEE.
22. Gettman, M.T., Kondraske, G.V., Traxer, O., Ogan, K., Napper, C., Jones, D.B., Pearle, M.S. and Cadeddu, J.A., 2003. Assessment of basic human performance resources predicts operative performance of laparoscopic surgery. Journal of the American College of Surgeons, 197(3), pp.489-496.
23. Dillon, W.E., Kondraske, G.V., Everett, L.J. and Volz, R.A., 2000. Performance theory based outcome measurement in engineering education and training. IEEE Transactions on Education, 43(2), pp.92-99.
24. https://www.nationsreportcard.gov/
25. Organisation for Economic Co-Operation and Development "OECD". 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. ISBN 978-92-64-04345-9.
26. United Nations Development Programme "UNDP". 2018. Human Development Indices and Indicators, 2018 Statistical Update. New York, NY.
27. Khalaf, K.A., Parnianpour, M. and Karakostas, T., 2000. Surface responses of maximum isokinetic ankle torque generation capability. Journal of Applied Biomechanics, 16(1), pp.52-59.

28. Masliah, M.R. and Milgram, P., 2000, April. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 25-32).
29. Tofallis, C., 2014. Add or multiply? A tutorial on ranking and choosing with multiple criteria. INFORMS Transactions on education, 14(3), pp.109-119.
30. Sarker, P.S., Sadanandan, S.K. and Srivastava, A.K., 2022. Resiliency Metrics for Monitoring and Analysis of Cyber-Power Distribution System with IoTs. IEEE Internet of Things Journal.
31. Bresolin Goncalves, A.P., 2022. The linguistic and cognitive mechanisms underlying language tests in healthy adults: a principal component analysis. M.S. Thesis, University of Montreal.
32. Morse, W., 1947. The flight envelope. Aeronautics. 17, 42-44.

GEORGE V. KONDRASKE
Dr. Kondraske serves as Professor of Electrical Engineering and Bioengineering at the University of Texas at Arlington. His research focuses on modeling and measurement of human performance. Over the past 40 years he has developed systems performance theory and various human performance measurement systems. These are employed in lab, clinic, and web-based/mobile contexts with applications ranging from Parkinson's Disease and rehabilitation engineering to academic, surgeon, and sport performance. He is a Fellow of the IEEE and a charter Fellow of the National Academy of Inventors.

LAWRENCE R. DISALVI
Mr. DiSalvi is currently pursuing a doctoral degree in Bioengineering at the University of Texas at Arlington. A former military helicopter pilot, his career includes 25 years in the aerospace industry as an engineering development and test director. His research interests include human performance modeling and human factors engineering.