Predicting Academic Performance in Surgical Training

Michael J. Yost, PhD,[*] Jeffery Gardner, MS,[*] Richard McMurtry Bell, MD,[†] Stephen A. Fann, MD,[*] John R. Lisk, BA,[‡] for the TriMetrix and Success Research Group, William G. Cheadle, MD,[§] Mitchell H. Goldman, MD,[‖] Susan Rawn, RN,[‖] John A. Weigelt, MD,[¶] Paula M. Termuhlen, MD,[¶] Randy J. Woods, MD,[#] Erick D. Endean, MD,[**] Joy Kimbrough, RN,[**] and Michael Hulme, MD[††]

[*]Department of Surgery, Medical University of South Carolina, Charleston, South Carolina; [†]Department of Surgery, University of South Carolina School of Medicine, Columbia, South Carolina; [‡]Lisk Associates, Lexington, Kentucky; [§]Department of Surgery, University of Louisville School of Medicine, Louisville, Kentucky; [‖]Department of Surgery, University of Tennessee School of Medicine, Knoxville, Tennessee; [¶]Department of Surgery, Medical College of Wisconsin, Milwaukee, Wisconsin; [#]Department of Surgery, Boonshoft School of Medicine, Wright State University, Dayton, Ohio; [**]Department of Surgery, University of Kentucky School of Medicine, Lexington, Kentucky; and [††]Department of Surgery, Bowman Grey School of Medicine, Wake Forest University, Winston-Salem, North Carolina

INTRODUCTION: During surgical residency, trainees are expected to master all the 6 competencies specified by the ACGME. Surgical training programs are also evaluated, in part, by the residency review committee based on the percentage of graduates of the program who successfully complete the qualifying examination and the certification examination of the American Board of Surgery in the first attempt. Many program directors (PDs) use the American Board of Surgery In-Training Examination (ABSITE) as an indicator of future performance on the qualifying examination. Failure to meet an individual program's standard may result in remediation or a delay in promotion to the next level of training. Remediation is expensive in terms of not only dollars but also resources, faculty time, and potential program disruptions. We embarked on an exploratory study to determine if residents who might be at risk for substandard performance on the ABSITE could be identified based on the individual resident's behavior and motivational characteristics. If such were possible, then PDs would have the opportunity to be proactive in developing a curriculum tailored to an individual resident, providing a greater opportunity for success in meeting the program's standards.

METHODS: Overall, 7 surgical training programs agreed to participate in this initial study and residents were recruited to voluntarily participate. Each participant completed an online assessment that characterizes an individual's behavioral style, motivators, and Acumen Index. Residents completed the assessment using a code name assigned by each individual PD or their designee. Assessments and the residents' 2013 ABSITE scores were forwarded for analysis using only the code name, thus insuring anonymity. Residents were grouped into those who took the junior examination, senior examination, and pass/fail categories. A passing score of $\geq$70% correct was chosen a priori. Correlations were performed using logistic regression and data were also entered into a neural network (NN) to develop a model that would explain performance based on data obtained from the TriMetrix assessments.

RESULTS: A total of 117 residents' TriMetrix and ABSITE scores were available for analysis. They were divided into 2 groups of 64 senior residents and 53 junior residents. For each group, the pass/fail criteria for the ABSITE were set at 70 and greater as passing and 69 and lower as failing. Multiple logistic regression analysis was complete for pass/fail vs the TriMetrix assessments. For the senior data group, it was found that the parameter Theoretical correlates with pass rate ($p < 0.043$, $B = -0.513$, $\exp(B) = 0.599$), which means increasing theoretical scores yields a decreasing likelihood of passing in the examination. For the junior data, the parameter Internal Role

The study concept was given by Bell, Yost, Lisk, and Fann. The members of the TriMetrix and Success Research Group helped in data collection and submission. Yost and Gardner did data analysis and interpretation of data. The manuscript was drafted by Bell. Manuscript revisions and intellectual contributions were by Bell, Yost, Fann, Gardner, Lisk, Cheadle, and Woods.

This research was supported, in part, by the Bonnstetter Family Foundation, providing the TriMetrix online assessments.

Correspondence: Inquiries to Richard M. Bell, MD, FACS, Department of Surgery, USC, 2 Medical Park, Suite 306, Columbia, SC 29203; fax: (803) 434-7349; e-mail: Richard.Bell@uscmed.sc.edu

  1931-7204/$30.00 http://dx.doi.org/10.1016/j.jsurg.2014.11.013

Awareness correlated with pass/fail rate ($p < 0.004$, $B = 0.66$, $\exp(B) = 1.935$), which means that an increasing Internal Role Awareness score increases the likelihood of a passing score.

The NN was able to be trained to predict ABSITE performance with surprising accuracy for both junior and senior residents.

CONCLUSION: Behavioral, motivational, and acumen characteristics can be useful to identify residents "at risk" for substandard performance on the ABSITE. Armed with this information, PDs have the opportunity to intervene proactively to offer these residents a greater chance for success. The NN was capable of developing a model that explained performance on the examination for both the junior and the senior examinations. Subsequent testing is needed to determine if the NN is a good predictive tool for performance on this examination. (J Surg 72:491-499. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: in-training examination, ABSITE, predicting academic success, TriMetrix, Medical Knowledge

COMPETENCIES: Interpersonal and Communication Skills, Professionalism, Medical Knowledge, Systems-Based Practice

INTRODUCTION

Medical Knowledge is 1 of the 6 competencies postgraduate surgical trainees are expected to master during the course of their residency.[1] Medical Knowledge is perhaps the most easily and objectively measured of the 6. Program directors (PDs) often use the American Board of Surgery (ABS) In-Training Examination (ABSITE) as an indicator of performance on the ABS Qualifying Examination (QE). PDs may base the need for remediation or promotion, at least in part, on the resident's performance in the ABSITE.[2] All PDs are aware that struggling residents require the additional expenditure of resources and can have disruptive consequences for the program in general.

The Residency Review Committee (RRC) for Surgery requires that 65% of program graduates successfully complete the QE and certifying examination (CE) of the ABS in the first attempt.[3] The RRC uses this standard as a quality indicator of the program's effectiveness in graduating residents with a sufficient knowledge base who are capable of practicing "without direct supervision."[4] Consequently, graduating residents' performance in the QE and CE potentially has a direct effect on the program's accreditation.

Residents bring a variety of personal talents to their training programs. It is presumed that residents possess the intellectual prowess to successfully master the cognitive demands of surgical training having obtained undergraduate degrees and advanced degrees by completing allopathic or osteopathic medical training. Academic performance is generally reflected by the residents' grade point average, but even stellar performance as an undergraduate does not always translate to successful academic performance as a postgraduate trainee.

We postulate that an assessment of behavioral, motivational, and acumen characteristics of individual residents can identify a group of trainees who might be at risk for substandard academic performance. Our initial experience with the TriMetrix tool suggested this to be the case. These characteristics are generally not identified through the general application process or the process for selecting residents for training positions. If such a group could be identified, this would allow prospective intervention to increase the chances for successful academic performance. This might include the construction of specific curricula tailored for each person, individualized mentoring, or developing learning strategies based on the trainees' characteristics. Such a program might obviate the need for remediation.

MATERIALS AND METHODS

We invited 10 surgical programs to participate in this exploratory first step, of them, 7 programs accepted. "Exempt" status approval for the project was granted by each individual participating program's institutional review board. Participation by the residents was purely voluntary, and each resident signed a consent to participate. Anonymity was ensured by having each PD or an institutional designee assign a random code name to be used by each resident to complete the assessment and for submission of the assessment results and ABSITE scores. PDs also indicated which participants took the junior or the senior examinations.

The online assessments were provided by Target Training International, Ltd, Phoenix, Arizona. The specifics of the TriMetrix assessments have been previously described in detail.[5,6] Briefly, the TriMetrix assessment consists of 3 parts. First, DISC, which is an assessment of behavioral style: D = Dominance, or how an individual deals with problems; I = Influencing, or how an individual deals with people; S = Steadiness, or how one deals with the pace of his or her job; and C = Compliance, or how one deals with rules, regulations, and boundaries. The assessment provides

TABLE 1. Resident Participation From Each of the 7 Training Programs Along With the Percentage Participation From Each Program. Analysis After Removal of Program G From the Data Set had No Influence on the Results

Program	Resident Participants
A	19 (100%)
B	22 (82%)
C	35 (76%)
D	15 (38%)
E	10 (31%)
F	14 (29%)
G	2 (6%)
Total	117 (48%)

TABLE 2. Summary Data for DISC and ABSITE Performance by Junior/Senior Examinations. The Following Data are not Continuous (Discrete Ordinal)

	Senior (64)		Junior (53)	
	Pass	**Fail**	**Pass**	**Fail**
$D_{(a)}$	38.2 ± 20.9	41.0 ± 23.0	34.7 ± 20.4	40.3 ± 26.1
$D_{(n)}$	42.4 ± 19.7	41.9 ± 29.4	37.7 ± 20.2	45.4 ± 23.1
$I_{(a)}$	47.7 ± 28.6	53.0 ± 29.3	49.8 ± 23.0	48.2 ± 26.3
$I_{(n)}$	50.3 ± 28.6	50.7 ± 30.0	52.8 ± 22.3	57.8 ± 24.2
$S_{(a)}$	60.5 ± 21.4	55.7 ± 24.8	63.4 ± 21.1	56.5 ± 24.6
$S_{(n)}$	67.9 ± 18.1	62.4 ± 18.0	67.8 ± 20.4	59.2 ± 28.0
$C_{(a)}$	48.8 ± 25.2	65.0 ± 24.7	62.0 ± 20.6	64.5 ±16.8
$C_{(n)}$	56.1 ± 26.0	61.1 ± 28.0	54.6 ± 20.6	53.9 ± 21.3

$D_{(a)}$, Dominance—adapted; $D_{(n)}$, Dominance—natural; $I_{(a)}$, Influenceing—adapted; $I_{(n)}$ = Influencing—natural; $S_{(a)}$, Steadiness—adapted; $S_{(n)}$, Steadiness—natural; $C_{(a)}$, Compliance—adapted; $C_{(n)}$ = Compliance—natural.
Mean and standard deviation.

insight into an individual's natural and adapted style (or how one must adjust one's behavior at work). This is annotated as $D_{(n)}$ or $D_{(a)}$. Most individuals display each of these behaviors to varying degrees, but usually one characteristic is dominant (Table 2). DISC characteristics are the "how" of behavior, developed by Marston,[7] a prominent American psychologist and inventor, in 1921. Spranger,[8] a German psychologist, developed a motivational analysis addressing the "why" of behavior, which is the second component of the analysis. It is an assessment of what motivates an individual to do his or her job. We accessed 6 primary motivators. Theoretical (THEO) reflects a search for truth or knowledge. Utilitarian reflects a desire for return on investment of time, effort, or resources. Aesthetic is a desire for form, beauty, and harmony. Social is a desire to help others, even at the expense of self. Individualistic refers to the desire to be in control of one's own destiny and the destiny of others. Traditional reflects the need to live guided by a code of conduct: religious, social, or ethnic. As with DISC, most people demonstrate each of the individual motivators to some degree but one is usually primary (Table 3). The third component of the TriMetrix is the Hartman Value Profile (HVP) or Acumen Index/Dimensional Balance developed by Hartman.[9] It is based on the theory of axiology, what an individual values, and the process of determining the value. It examines how an individual views the external world regarding other people, practical thinking, and systems judgment. The assessment also provides insight into an individual's view of oneself regarding one's sense of self, role awareness, and the future. It is the most flexible of the 3 assessments and can change based on an individual's current situation (Table 4).

The TriMetrix provides 20 separate data points: 8 for DISC, 6 for motivators, and 6 for the HVP. These data points were analyzed by logistic regression and used to train the neural network (NN). Correlations were determined with the ABSITE score, and an initial model was constructed to explain ABSITE performance. As the ABS offered 2 versions of the ABSITE in 2013, a junior and a senior examination, the residents were divided into these 2 groups for analysis. A score of 70% correct was chosen as a "pass/fail" cut point because the ABS has historically suggested that performance at the 30 percentile (generally equivalent to 70% correct responses) in the final year of surgical training provides some assurance that the resident has a good chance of passing the QE in the first attempt.

An example of individual data generated by the TriMetrix assessment is shown in Figures 1 to 3, which graphically

TABLE 3. Summary Data for Motivators and ABSITE Performance by Junior/Senior Examinations

	Senior (64)		Junior (53)	
	Pass	**Fail**	**Pass**	**Fail**
n	57 (0.89)*	7 (0.11)*	37 (0.73)*	16 (0.27)*
Mean	76.7	57.9	75.0	51.9
σ	4.5	7.2	8.3	7.6
Theoretical (THEO)	50.7 ± 8.9	57.9 ±7.2	51.4 ± 8.3	57.9 ± 7.6
Utilitarian (UTIL)	40.7 ± 11.1	36.3 ± 10.0	44.3 ± 12.0	44.3 ± 13.9
Aesthetic (AEST)	33.5 ± 10.1	27.9 ± 9.8	32.6 ± 10.4	34.7 ± 11.6
Social (SOC)	45.8 ± 9.3	53.4 ± 11.1	46.4 ± 10.2	46.0 ± 11.2
Individualistic (IND)	41.4 ± 8.6	40.3 ± 4.2	38.7 ± 7.9	40.4 ± 7.1
Traditional (TRAD)	37.1 ± 9.3	36.3 ± 13.3	38.6 ± 10.4	35.9 ± 9.7

*Number, percentage.

TABLE 4. Summary Data for HVP and ABSITE Performance by Junior/Senior Examinations. The Following Data are not Continuous (Discrete Ordinal)

	Senior (64)		Junior (53)	
	Pass	**Fail**	**Pass**	**Fail**
Ex_UO	8.2 ± 1.2	8.6 ± 0.7	8.0 ± 1.3	7.9 ± 1.4
Ex_PT	8.2 ± 1.1	8.2 ± 0.4	8.1 ± 1.3	7.9 ± 1.4
Ex_SJ	7.8 ± 1.2	7.7 ± 1.0	7.8 ± 1.2	7.4 ± 1.3
Io_SS	6.8 ± 1.5	7.9 ± 1.1	6.8 ± 1.5	6.4 ± 1.5
Io_RA	6.7 ± 1.5	7.7 ± 0.9	6.7 ± 1.4	5.6 ± 1.4
Io_SD	6.9 ± 1.2	7.5 ± 0.7	7.0 ± 1.2	6.8 ± 1.1

Ex, external (world view); IO, internal (self-view); PT, practical thinking; RA, role awareness; SD, self-direction; SJ, systems judgment; SS, sense of self; UO, understanding others/empathetic outlook.
Mean and standard deviation.
UO = Measures how well an individual understands other people and has an empathetic outlook.
PT = Reflects how clearly an individual sees how to obtain results or solve problems.
SJ = Provivdes an indication of how clearly an individual sees the "big picture" or how the system operates.
SS = Reflects an individual's comfort level with who he or she is.
RA = Measures the clarity that an individual sees his or her roles in life, both professionally and personally.
SD = Gives insight as to how clearly an individual sees his or her future.
The Hartman Value Profile provides only a reflection of these parameters at the time the assessment was taken. It is subject to change based on life and work circumstances and as such is the most mutable of all the variables collected.

show the results of DISC, motivators, and the HVP/Acumen Index. Summary information provided by Target Training International Ltd is also shown from the general population as of 2011 as well as the coefficient of reliability. Summary data for the HVP are not provided, as this assessment changes based on the individual's perceptions of the situation in which he or she currently finds himself or herself. Data do not exist for surgeons in training as a specific group at present.

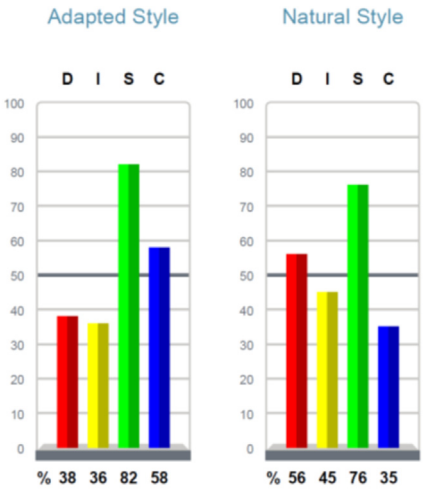
To analyze the data, 2 separate techniques were used. First, we used binary logistic regression to analyze the pass/fail rate for residents taking either the senior or the junior examinations. Second, we used an artificial NN to attempt to model actual scores based on combinations of the TriMetrix parameters.

Binary Logistic Regression

Data analysis was undertaken by first screening for correlations between the variables measured with TriMetrix and ABSITE test scores. Correlation coefficients did not pass 95% confidence, which is normally used as the cutoff for statistical significance. Several did pass the 90% confidence test, which indicates there could be important information in the data but that greater numbers of residents may be need to gain traditional statistical confidence. To best communicate the results of this study, we used 2 methods of analysis. First, we determined if we could predict residents who may not achieve a "passing" ABSITE score and thus allow a more proactive approach in structured study guidance during the training year. For this, we used a binary logistic regression. Logistic regression is used to predict a categorical variable from a set of predictor variables. Pass/fail on the ABSITE was assigned as pass a value of 1 and fail a value of 0. The regression is performed and a likelihood function is generated. Likelihood or probability is continuous between 0 and 1.

Neural Network

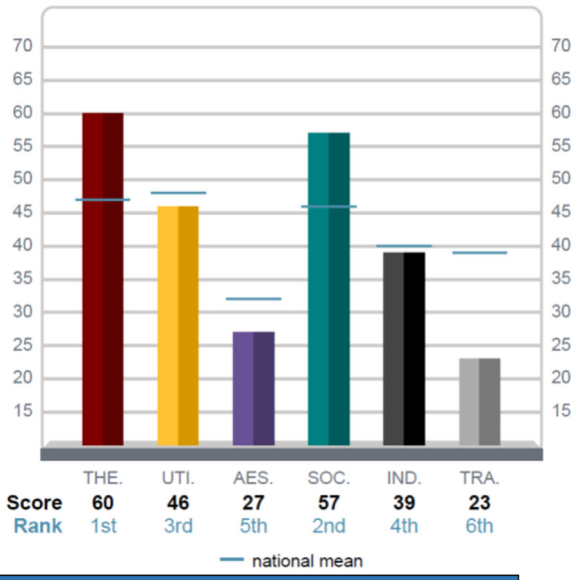
A NN is a function, given training and testing data that can be used to predict an output when given a new input. Originally designed to crudely mimic the connectivity in a human brain, it functions by strengthening or weakening



Distribution of scores from the General Population, n = 17,801	
Behavior	**Mean ± Standard Deviation**
Dominance	45.56 ± 16.39
Influence	60.29 ± 15.37
Steadiness	54.74 ± 17.03
Compliance	46.81 ± 15.08

Cronbach's α for Behaviors in the General Population, n = 16,950								
	Adapted D	Natural D	Adapted I	Natural I	Adapted S	Natural S	Adapted C	Natural C
α	.885	.884	.850	.845	.856	.834	.826	.826

FIGURE 1. An example of a DISC (Behaviors) assessment along with data from the general population showing mean scores and standard deviation. The 50% "energy line" is used to determine the trainee's primary behavioral style. In this case, "steadiness" is farthest from the line for both the natural and the adapted behaviors. The results of the assessment consider not only the dominant behavior but also how other behaviors influence the way in which trainees respond to the challenges of their environment and their job. Reliability (Cronbach α) is excellent for this assessment. Data from medical trainees as a specific group are not available. (This graph was randomlly selected as an example from all residents who took the TriMetrix assessment. Means and standard deviations from the general population were provided by TTI Ltd., Phoenix AZ, Technical Documents, and is current as of 2011.)



Distribution of Scores from the General Population, n = 17,801	
Category	**Mean ± SD**
Theoretical	46.93 ± 9.37
Utilitarian	47.44 ± 10.49
Aesthetic	32.19 ± 9.88
Social	46.81 ± 9.91
Individualistic	39.96 ± 8.63
Traditional	38.66 ± 8.39

Cronbach's α for Motivators in the General Population, n =38,314	
Category	
Theoretical	0.755
Utilitarian	0.820
Aesthetic	0.822
Social	0.829
Individualistic	0.679
Traditional	0.705

FIGURE 2. Graphical representation of a trainee's motivational characteristics as well as general population norms, standard deviation, and the assessment of reliability of the evaluation. In this case, the primary motivator is Theoretical, or a desire for knowledge and truth. The social motivator is second highest, indicating a desire to use this knowledge for the benefit of others. Solid lines indicate mean values from the general population. (This graph was randomlly selected as an example from all residents who took the TriMetrix assessment. Means and standard deviations from the general population were provided by TTI Ltd., Phoenix AZ, Technical Documents, and is current as of 2011.)

connections between nodes or neurons using a mathematical algorithm. The NN attempts to "learn" the relationship between the predictor variables, the TriMetrix, and the outcome variable, the ABSITE score. NNs are robust pattern-matching algorithms that make no assumptions as to the shape or form of the relationship between the predictor data and the outcomes.

The data were divided into 2 groups: senior test scores and junior test scores. The first network model for each group consisted of all of the original variables. Each model was then trained using a predetermined split in the data consisting of 80% for training and 20% for testing purposes. Each network consisted of 20 input variables or nodes, a hidden layer of 20 nodes, and a single output node. The models were trained using the TriMetrix data for 30,000 iterations to minimize error and then tested on the unseen test data. The models were subsequently refined to identify the predictors that are more strongly associated with ABSITE scores from those that appear to have little or no association. We then used a gradient-based search technique; the slope of the function (predicted ABSITE score) for each of the variables (parameters from the online assessments) was calculated at scaled parameter values of 0.3, 0.5, and 0.75. Variables that had an absolute value of slope less than 0.005 were considered to be insignificant in the model. This resulted in the elimination of $S_{(n)}$ and Individualistic variables from the senior model and $I_{(a)}$, $S_{(a)}$, Traditional, and Internal Sense of Self variables from the junior model. The network models were retrained using the same specifications as before minus the eliminated input variables. The same gradient-based search was implemented again, identifying parameters that were significant and trying to eliminate other unnecessary variables. These gradient-based searches were repeated until the network did not train on the remaining set of predictors. We then revert to the last predictor set that successfully trained. Additional details concerning NN is provided in the Appendix.

RESULTS

Overall, 7 individual training programs agreed to participate. Of 242 potential participants, 117 (48%) completed the TriMetrix online assessment and had their ABSITE scores submitted for analysis. Removal of the 2 residents from Program G had no effect on the analysis (Table 1). Individual program resident participation varied from 6% to 100% (Table 1). The summary data are provided in Tables 2 to 4.

Results of the Binary Logistic Regression

Senior Data

A total of 117 residents took the examination, and ABSITE scores were submitted; 64 were seniors and 57 (89%) passed the examination. Multiple logistic regression analysis was completed for pass/fail vs the TriMetrix assessments. For the senior data group, it was found that the parameter "THEO" correlates with the pass rate ($p < 0.043$, $B = -0.513$, $\exp(B) = 0.599$), indicating that increasing theoretical scores yields a decreasing likelihood of a passing in the examination.

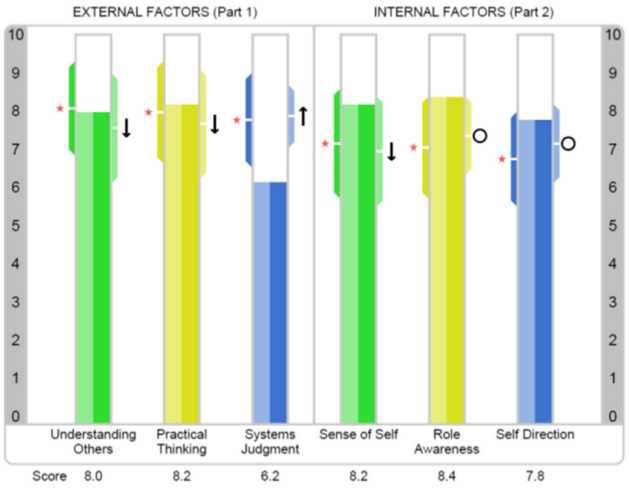


FIGURE 3. Graphical representation of the results of the Hartman Value Profile or Acumen Index/Dimensional Balance Assessment. It provides a snapshot of how the trainee sees the world and himself or herself in the world at the time the assessment was taken and as such is the most likely to change based on the current situation or circumstances. The mean population value is denoted by the red "star" on the left of each bar along with the brackets for the standard deviation. The up and down arrows and the circle represent "bias" for each individual component. The upward arrow implies overvaluation of each element, the downward arrow implies undervaluation, and the circle indicates that each of the respondents placed each of the dimensions in their correct postion or that there were and equal number of overvaluations and undervaluations. Bias has little importance to our particular study. (This graph was randomlly selected as an example from all residents who took the TriMetrix assessment. Means and standard deviations from the general population were provided by TTI Ltd., Phoenix AZ, Technical Documents, and is current as of 2011.)

Junior Data

Overall, 53 juniors took the examination and 37 (69.8%) passed. Multiple logistic regression analysis was complete for pass/fail vs the TriMetrix assessments. For the junior data, the parameter Role Awareness (IO_RA) correlated with pass/fail rate ($p < 0.004$, $B = 0.66$, $\exp(B) = 1.935$), indicating that an increasing IO_RA score increases the likelihood of a pass score.

NN Results

It was determined that the most significant variables driving each model positively were Influencing—adapted ($I_{(a)}$), THEO, and External—Empathetic Outlook; additionally, Influencing—natural ($I_{(n)}$) and External—Systems Judgment produced a negative influence for the senior model and Compliance—natural ($C_{(n)}$) a positive one. The trained network was then tested on the reserve data set and is reported graphically in Figure 4. The residual plots indicate that the variables identified have some influence on performance. Like most models, there are variables that contribute but have not been identified or included in the analysis. The NN and the logistic regression model are discordant regarding THEO for the senior examination: the NN shows a positive correlation and the regression model shows a negative correlation. This may seem contradictory, but considering that the NN is able to look at the influence of other variables, independently or in concert with multiple others, the results can be rationalized. In the study of behaviors and motivators, a high or low value of one may be modified by a high or low score in another or combinations of the others. Our hypothesis is that it is indeed an important variable, although we do not have enough data to understand its effect across all residents and possible scores. This is best seen in the results of the NN and the residual plots (Fig. 4). These plots of the residuals are expected and are because of many more unaccounted variables that certainly have an influence on the actual test scores. The scatter of the residuals suggests that the NN has identified a mathematical model that is coming close to predicting performance, and although not yet perfect, it shows promise.

DISCUSSION

This study is an initial exploration into the feasibility of developing a model to identify residents who might be considered potentially at risk for substandard academic performance using ABSITE scores as a surrogate marker. The premise is that certain behaviors, motivators, and acumen scores, singly or in combination with others, may provide this opportunity. Most PDs and surgical teaching faculty presume that surgical trainees have demonstrated their intellectual capacity to develop the required knowledge base by their previous undergraduate and medical school performance. Yet, there are a number of residents who struggle academically. Success as an undergraduate does not guarantee success as a postgraduate surgical trainee. The individual characteristics of residents described by the TriMetrix assessments are generally not apparent during the standard application and interview process conducted before determining the ranking order for a programs match list. It is our contention that behavioral, motivational, and acumen characteristics have an influence on academic performance.

The TriMetrix was chosen over the more familiar Myers-Briggs Type Indicator. First, Myers-Briggs Type Indicator is purely a personality assessment tool, whereas the TriMetrix evaluates behaviors, motivators, and a snapshot of how the individual sees the world and oneself in the world at the time the assessment was taken. Although there is an obvious overlap between behavior and personality, the value of the TriMetrix is that it provides data not only on how an individual responds to problems, people, pace, and procedures but also on why they do what they do. The TriMetrix has been extensively studied and validated in 40 languages and 20 countries. This assessment tool has excellent internal reliability (Figs. 1 and 2) and has been found free of any adverse effect when used as a part of selection processes in the business world. We have found that the TriMetrix

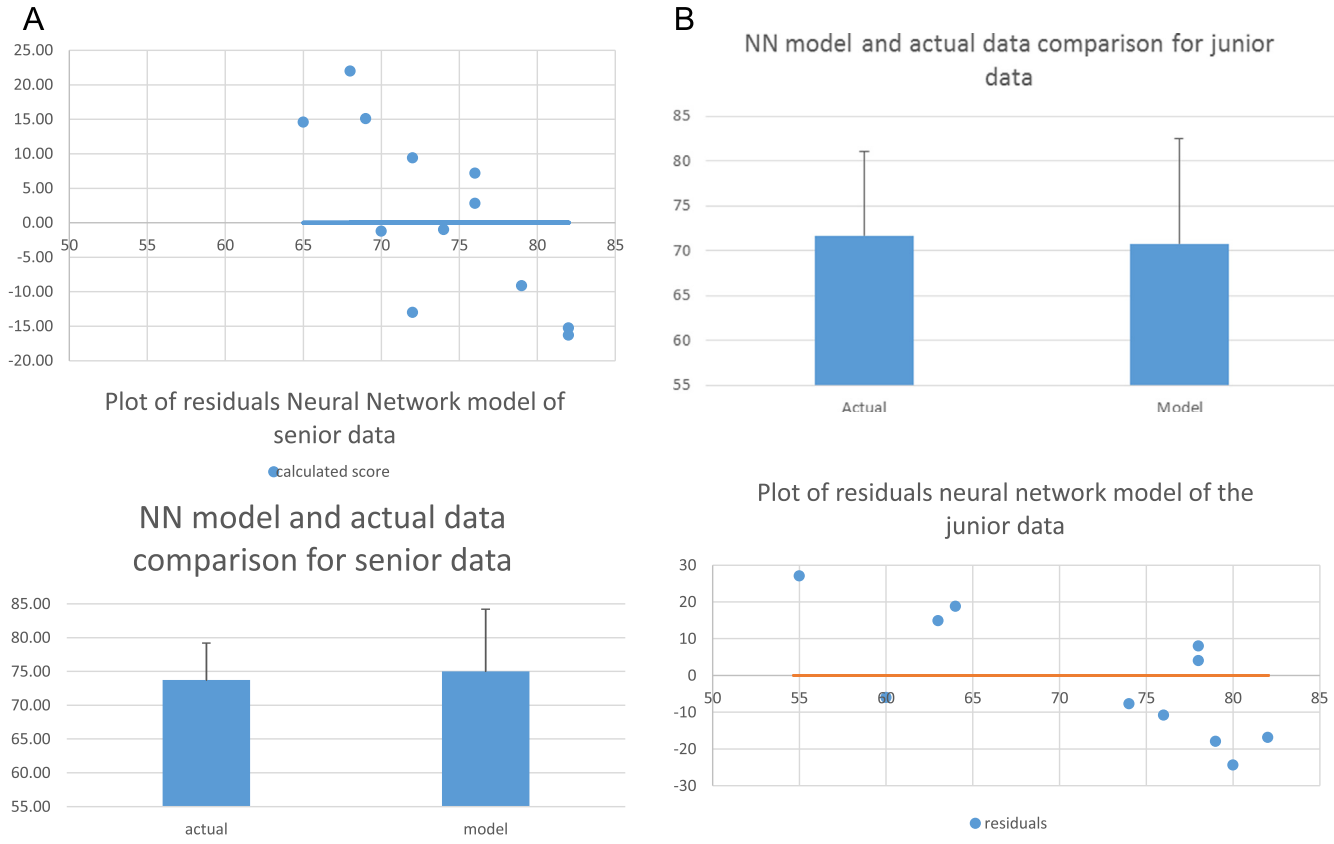


FIGURE 4. (A) A comparison of the actual raw score data with that predicted from the neural network model for test data previously unseen by the network during training. For the actual data, the mean is 73.75 ± 5.45. The model predicts a mean of 75.0 ± 9.2. Evaluation of the residuals indicates a trend. It appears the model predicts lower scores for those who actually scored high in the examination and higher scores than those who scored lower. This behavior indicates that, although some of the parameters tested are important, there remains missing parameters to accurately predict an individual's performance in the ABSITE (see section Discussion). Residuals represent the difference between the "test subject's" actual vs predicted ABSITE scores and the "actual" and "model" performance mean and standard deviation. (B) A comparison of the actual raw score data with that predicted from the neural network model for test data previously unseen by the network during training. For the actual data, the mean is 71.7 ± 9.4. The model predicts a mean of 70.8 ± 11.7. Evaluation of the residuals indicates a trend. It appears the model predicts lower scores for those who actually scored high in the examination and higher scores for those who scored lower. This behavior indicates that, although some of the parameters tested are important, there remains missing parameters to accurately predict an individual's performance on the ABSITE. However, from the mean and standard deviation, the model is able to better predict scores on the junior examination than on the senior examination. Residuals represent the difference between the "test subject's" actual vs predicted ABSITE scores and the "actual" and "model" performance mean and standard deviation.

allows the development of program benchmarks, which can be used to compare individual residents with the attributes of successful residents in the program. It is our contention that behavior and motivation contribute more to successful performance than personality alone does. The TriMetrix has not been extensively used in the medical field to date.

The QE of the ABS is a high-risk examination for graduate surgeons and now for the individual programs. The RRC has established a standard of 65% success for first-time takers of both the QE and the CE. PDs have little influence on the study habits and preparatory efforts of their residents after graduation. Hopefully, PDs and the program's curriculum have had a sufficient effect on the trainees to provide them with the greatest opportunity for success by allowing the trainee to develop the necessary knowledge base to pass the examination. It would seem intuitive that a tool that could identify potential academic difficulties would be extremely helpful for PDs. Attrition from surgical training remains a significant problem, from 3% to 26%.[10,11] Kelz et al.[12] identified poor ABSITE performance as a major factor in resident attrition. Every program has felt the pain of losing a resident for whatever reason: academic, clinical performance, personal issues, or a simple voluntary change in career goals. Although there are multiple reasons for residents altering their career choices,[13] a strategy that could help minimize the disruption caused by attrition due to substandard academic performance would be welcomed by all.

There is no disagreement that the performance on the first ABSITE is an indicator of a struggling resident, at least academically, and as such might save the expense and effort of assessments such as the TriMetrix. Furthermore, an analysis of learning styles, i.e., auditory, visual, and psychomotor, can provide useful information to assist in the development of an individualized curriculum and study plans. Although this is certainly useful information, learning

style alone does not provide insight into the how and why of performance.

Juniors with a high clarity score (IO_RA) correlated with a passing score on the ABSITE. This is not surprising, as those who have a strong sense of their professional and personal roles in life would be expected to do well. What did surprise is that the higher THEO had a negative correlation with the ABSITE performance for the senior examination by binary logistic regression, which is initially counterintuitive. It is impossible to develop a complete explanatory model using the NN, as all possible contributing factors have not been included. We considered the possibility that residents who have an extensive knowledge base but performed below prediction might have known "too much" about a question or "overanalyzed" their responses. As we have no evidence to support this position, we favor our impression that the complexity of the NN and the extremes of the variables, e.g., THEO, are probably responsible.

The models developed by logistic regression and the NN do not coincide completely. This is not surprising considering the differences in the methodology. The NN is far more complex than logistic regression is and evaluates the influence of variables on the dependent variable, and each other, to a much greater degree, from thousands of interactions and combinations. Furthermore, the NN does not assume the shape (positive vs negative) of the function. What is interesting is that both approaches identified THEO as an influence; it may be the extremes of this variable that accounts for the difference in sign. We propose this as the explanation for the discordance between binary logistic regression and the NN in the influence of THEO in the senior group analysis. It is our feeling that the NN is much more robust in developing a mathematical model, and although not accurate at predicting ABSITE scores in this exploratory study, it is beginning to identify factors that contribute to performance.

Our study has a number of limitations. We are under no illusion that our assessment identifies all the factors that influence performance on an isolated examination. Fatigue, mood, emotions, preparatory effort, illness, and many other factors are operative. In addition, our study does not address test-taking skills that may have just as much an effect on performance as the breadth of knowledge a resident brings to the standardized examination. The relative low number of resident participants (48%) and the disparity in the levels of participation by each of the programs may skew our results. Selection bias is certainly a possibility. Furthermore, our study population represents only a fraction of all general surgical trainees in the United States. It is apparent that these results need to be validated in a much larger prospective study. Some programs included preliminary residents, and this may have influenced the results from the perspective that these residents may have a degree of uncertainty regarding their future or how much they had at stake in their ABSITE performance. As all programs are US based, extrapolation of these findings to programs in other countries is not appropriate as training varies greatly throughout the world. The data are problematic as they are not continuous and do not follow a normal distribution. Consequently, we chose to use logistic regression and the NN. NNs have been criticized as artificially developing an explanatory mathematical model when one does not actually exist in the real world. The findings of the NN need to be confirmed with a prospective look at its predictive accuracy. The study could also be criticized for its potential lead-time bias, having chosen only surgical residents who have already selected a career based on their "fit" into the surgical arena or their specific training program. Although there are certain characteristics of the residents who participated in the study that are similar, there is sufficient diversity to exclude the possibility that the choice of a surgical career is predetermined by these characteristics. Finally, the ABS chose to move to a single ABSITE examination in 2014; hence, it remains to be determined if these results can be extrapolated to subsequent iterations of the examination.

CONCLUSIONS

This preliminary study suggests that behavioral and motivational characteristics of surgical trainees can potentially be used to identify residents who might be at risk for substandard academic performance. Our results suggest that these characteristics, when entered into a NN, allow the development of a mathematical model that is coming close to predicting academic performance on the in-training examination, and although not yet perfect, it shows promise. More importantly, this model may help identify residents at risk for substandard academic performance and provide the opportunity for PDs to be proactive in developing strategies to improve the residents' chances for academic success. Expanding the study population will allow validation of the model. The effect of the change introduced in 2014 by the ABS by returning to a single examination for both juniors and seniors is unknown.

REFERENCES

1. ACGME Program Requirements for Graduate Medical Education in General Surgery, IV.5.A.b. Available at: ⟨http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/440_general_surgery_01012008_07012012.pdf⟩. Accessed 23.4.14.
2. ACGME Program Requirements for Graduate Medical Education in General Surgery, V.A.1.e. Available at: ⟨http://www.acgme.org/acgmeweb/Portals/0/PFAssets/

ProgramRequirements/440_general_surgery_01012008_07012012.pdf⟩. Accessed 23.4.14.

3. ACGME Program Requirements for Graduate Medical Education in General Surgery, V.C.3. Available at: ⟨http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/440_general_surgery_01012008_07012012.pdf⟩. Accessed 23.4.14.

4. ACGME Program Requirements for Graduate Medical Education in General Surgery, V.A.2.b. Available at: ⟨http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/440_general_surgery_01012008_07012012.pdf⟩. Accessed 23.4.14.

5. Bell RM, Fann SA, Morrison JE, Lisk JR. Determining the personal talents and behavioral styles of applicants for surgical training: a new look at an old problem, Part I. *J Surg Educ*. 2011;68(6):534-541.

6. Bell RM, Fann SA, Morrison JE, Lisk JR. Determining the personal talents and behavioral styles of applicants for surgical training: a new look at an old problem, Part II. *J Surg Educ*. 2012;69(1):23-29.

7. Marston WM. The Emotions of Normal People. New York: Harcourt, Brace & Co; 1928.

8. Spranger E. Types of Men: The Psychology and Ethics of Personality. [translation by Pigors PJW, Trans.] New York: G. E. Stechert Company; 1928 [Original work by Lebensformen; Halle (Saale): Niemeyer, 1914].

9. Hartman RS. The Structure of Value: Foundations of Scientific Axiology. Carbondale: Southern Illinois University Press; 1967.

10. Morris JB, Leibrant TJ, Rhodes RS. Voluntary changes in surgical career paths: a survey of program directors in surgery. *J Am Coll Surg*. 2003;196(4):611-616.

11. Yeo H, Viola K, Berg D, et al. Attitudes, training experience and professional expectations of US general surgery residents: a national survey. *J Am Med Assoc*. 2009;302(11):1301-1308.

12. Kelz RR, Mullen JL, Kaiser LR. Prevention of surgical resident attrition by a novel selection process. *Ann Surg*. 2010;252(3):537-543.

13. Sullivan MC, Yeo H, Roman SA, et al. Surgical residency and attrition: defining the individual and programmatic factors predictive of trainee loss. *J Am Coll Surg*. 2013;216(3):461-471.

SUPPLEMENTARY MATERIALS

Supplementary material cited in this article is available online at doi:10.1016/j.jsurg.2014.11.013.