An Application of Logistic Regression in Identifying Target Populations Using TriMetrixEQ® Variables

Eric Gehrig, PhD
Senior Research Scientist
Research & Development
Target Training International

Robert Stokes
Research Assistant
Research & Development
Target Training International

September 26, 2019

Abstract

The APA's position on evidence of validity includes a category called "relationships to other variables". This concept appears to be an updated version of what was once called criterion validity and was comprised of two further validity requirements, concurrent and predictive. The study below establishes relationships between several external variables based on demographic information obtained using the O*Net job classification model and the scales of the TTI Success Insights TriMetrixEQ® assessment. The work uses a logistic regression modeling approach to derive statistically significant functional relationships between the TriMetrixEQ scales and membership in the job classification group of interest. ROC curve analysis is used to show the classification algorithm outperforms the standard random selection technique.

Introduction

The American Psychological Association (APA) identifies five key areas of any psychometric assessment requiring evidence of validity, see [16]. One of these areas is *Relationship to Other Variables* which encompasses several concepts such as criterion, concurrent, and predictive validity. While [9] presents many of the theoretical concepts of evidence of reliability and validity of assessments, [2] is an effort by the Joint Committee on Standards for Educational and Psychological Testing to describe in more detail guidelines related to these ideas.

As an example, Standard 1.4 of [2] states

> If a test score is interpreted for a given use in a way that has not been validated, it is incumbent on the user to justify the new interpretation for that use, providing a rationale and collecting new evidence, if necessary.

There are at least two major takeaways from the previous statement. The first is more implicit in that any use of assessment data in a way that has not been validated requires validation. Second and more explicit is that the responsibility for providing evidence of such validation lies primarily with the test user.

One of the purposes of this paper is to begin to establish a foundation of evidence of validity for the TTI Success Insights TriMetrixEQ assessment by combining demographic information, specifically job description or classification, with a classification algorithm. In this case, the authors choose to use a logistic regression approach to assigning membership in a group.

In a previous study, [8] shows that the TTI TrimetrixDNA Legacy® variables may be used in conjunction with a logistic regression classification approach to successfully differentiate serial entrepreneurs from a group randomly selected from the TTI database. This study is an example demonstrating the relationship between TTI assessment derived variables (Behaviors, Motivators, and DNA) with other variables (membership in a specific sub-population within a larger population).

In April, 2018, TTI began offering the opportunity for our Network of Value Added Associates (VAA) to add demographics collection to their assessment links. Since that time a good deal of demographic information has been collected and analyzed in various ways. The current study under consideration asks whether the use of a classification technique in a manner similar to the previously discussed paper, see [8], may be applied to subsets of the demographic data. In particular, this study considers several subsets defined through the use of the Job Description demographic category in the TTI demographics database.

This paper is organized as follows. The next two sections discuss some details of the data sets under consideration including an internal consistency presentation. This is followed by a section briefly discussing some of the details of logistic regression. A section discussing measurement of the success or failure of the model is presented next, followed by a section presenting the results of the study on several data sets. Finally, a summary and future work section wraps up the paper.

TriMetrixEQ Data and the Appropriate Subsets

The TTI Success Insights TriMetreixEQ assessment is comprised of three separate assessments. The first, Style Insights®, is used to measure observable behavior and is based on the four factor DISC model. The four factors stand for **D**ominance, **I**nfluence, **S**teadiness, and **C**ompliance. The second is called Motivation Insights® and is based on a six factor model and is used to measure the inherent motivation behind an individual's actions. The scales measured are Theoretical, Utilitarian, Aesthetic, Social, Individualistic, and Traditional. The third assessment is the Emotional Quotient (EQ) assessment that measures an individual's emotional intelligence. The scales measured are Self Awareness, Self Regulation, Social Awareness, Social Regulation, and Motivation.

An additional set of variables is considered during the analysis below. There are a set of 12 variables derived from the DISC variables of the Style Insights assessment. They are jointly called the Behavioral Hierarchy. This set contains the variables Urgency, Frequent Interaction with Others, Organized Workplace, Analysis of Data, Competitiveness, Versatility, People Oriented, Frequent Change, Customer Relations, Follow Up and Follow Through, and Following Policy.

Additionally, Motivation Insights generates a total of 12 separate scores that are collectively known as Driving Forces. In the section containing the results below, the regressions consider the 12 Driving Forces variables while analysis of the scales focuses on the 6 scales mentioned above. There is no inconsistency in this approach. All information is generated by the same assessment taker responses.

All data considered in this study has been collected between April 2, 2018 and September 19, 2019. The data sets are pulled from the TTI Internet Delivery System databases. A data set is generated for each of the three assessments, along with a combined data set generated by individuals taking the TriMetrixEQ combined assessment.

Each of the data sets comes with a series of demographic data categories that may or may not be utilized in any given instance. For those data entries that contain non-empty demographic information, a brief study of categorical breakdowns reveals a series of subsets of the larger data sets that may be individually studied. Of primary interest for this study is the Job Description category which contains a list of job descriptions as defined by O*NET. For more information on O*NET one may consult the United States Department of Labor Employment and Training Administration O*NET website at `https://www.doleta.gov/programs/onet/eta_default.cfm`.

Several job classifications are clearly chosen more frequently than others making them particularly important to TTI Success Insights given their higher frequency in the current data sets.

Table 1: Most Commonly Chosen Job Descriptions: TriMetrixEQ

Title	Frequency
Sales Rep, Services, All Other	200
Managers, All Other	127
Accountants	107
CEO	107

The highest frequency job descriptions chosen in the combined TriMetrix data set are summarized in Table 1.

Ideally, it is preferable to have larger subsets of data to study. However, it is the opinion of these authors that the frequencies presented in Table 1 are large enough to present an initial look at these categories with an eye toward the future when much more data becomes available. The purpose of this initial look is to establish relationships between TTI Success Insights assessment variables and external variables, specifically Job Description. Future studies are planned for such a time when more data are available and, preferably, access to other criteria such as a performance metric are also available.

The data sets used in the study are then formed by taking the set of information of interest and combining them with a randomly generated sample of an appropriate size. For this study, each set of interest is combined with a random sample of size $2.5 * n$, where n is the number of data points in the set in question. No selection criteria are used in addition to the requirement that the random subset comes from the larger TriMetrixEQ data set and that the random subset does not contain any individuals from the population of interest.

We shall use the naming convention as outlined in Table 2.

Table 2: Naming Convention

Title	Data Set
Sales Rep, Services, All Other	Sales1
Managers, All Other	Managers
Accountants	Accountants
CEO	CEO

Before discussing some of the statistics related to measures of evidence of reliability and validity, we present some basic information obtained from the demographics for completeness.

Table 3: Basic Information Sales1

Category	N	Avg Age	M/F Ratio
Sales1	200	41	120/80
Managers	127	48	71/56
Accountants	107	45	41/66
CEO	107	55	81/26

These are the top job classifications in the TTI Success Insights database by number of respondents. Given that the main focus is a comparison of the current data to that of the previously mentioned study on Serial Entrepreneurs, see [8], a decision was made to concentrate on the CEO data from TriMetrixEQ with an eye on future publications with additional information as it becomes available.

Internal Consistency Estimates for TriMetrix Data Subsets

Internal consistency estimates help assess the consistency of the individual scales of an assessment. There are several measures one may use to assess internal consistency. For purposes of this exposition, the authors choose to present the most commonly reported measure, the α coefficient originally derived in [13],with further presentations in [10] and [12], and, finally, made famous by the work in [4].

The tables below also present two measures related to evidence of reliability and validity. The first is the average inter item correlation which is directly proportional to the value of the α coefficient. The second is the average corrected item total correlation. Corrected item total is a measure that may be interpreted to some extent as a proxy for internal structure evidence of validity, see [14].

Table 4: Common Reliability Coefficient Interpretation, see [5]

Range	Interpretation
$\alpha \geq 0.90$	Shorten Scale
$0.80 \leq \alpha < 0.90$	Very Good
$0.70 \leq \alpha < 0.80$	Respectable
$0.65 \leq \alpha < 0.70$	Undesirable
$\alpha < 0.65$	Unacceptable

While there is not a consensus on what specifically are acceptable levels of average inter item correlation and average corrected item total correlation, we cite [3] for a source on ranges to guide the discussion. For inter item correlation, the authors of [3] state two generally acceptable ranges that depend on the intended scope of the scale. The first range is 0.20 to 0.40 for a narrower scope and 0.15 to 0.50 for more encompassing scale. Generally speaking, the desired range for corrected item total correlation is 0.30 to 0.50. The reader should note that these are ranges and guidelines, not hard cutoff points.

Whether a measure is acceptable or not is often a matter of interpretation of the scale, its intended purpose, and many other variables. A general rule of thumb is that items with very low item correlation provide little information and those with too high become redundant. The respective averages provide a summation of the overall effectiveness of the scale.

Finally, there is a more sound foundation for interpreting the internal consistency measure, although there are differing opinions in this arena as well. However, for the purposes of this study, the authors use the guidelines presented in Table 4.

Table 5 presents the α coefficient, average inter-item correlation, and average corrected item to total score correlation. The alpha coefficient is a proxy for the internal consistency measure of an assessment scale. The average inter-item correlation speaks to the evidence of reliability and, to a certain extent, the homogeneity of the scale. The concept of homogeneity of the scale is tied to the ideas of evidence of validity. Finally, the average corrected item to total correlation is another proxy for evidence of validity.

Table 5: Behavior Statistics: CEO

Scale	Std α	Avg IIC	Avg ITC
D	0.89	0.26	0.48
I	0.87	0.22	0.44
S	0.84	0.18	0.38
C	0.84	0.18	0.40

The overall performance of the DISC variables from the Style Insights assessment shows strong measures of internal consistency and the two correlation statistics, especially considering the small subset of data under consideration. Given that we are providing a study based on generating evidence of validity, in particular, evidence of validity based on relationships to external variables (predictive validity), it is highly important to first establish evidence of reliability.

Table 6: Motivators Statistics: CEO

Scale	Std α	Avg IIC	Avg ITC
The	0.88	0.37	0.57
Uti	0.84	0.30	0.50
Aes	0.76	0.21	0.40
Soc	0.87	0.38	0.57
Ind	0.78	0.23	0.42
Tra	0.84	0.31	0.50

The information presented in Table 6 shows that the Motivators data performs in a similar fashion to the Behaviors data shown in Table 5. As noted previously, a precursor to establishing evidence of validity is to have evidence of reliability. This analysis shows that there is a reasonable amount of internal consistency and homogeneity in the Motivators scales.

Table 7: Emotional Quotient Statistics: CEO

Scale	Std α	Avg IIC	Avg ITC
SA	0.83	0.33	0.52
SR	0.82	0.27	0.47
M	0.83	0.29	0.48
SoA	0.81	0.26	0.46
SoR	0.79	0.26	0.45

The data presented in Table 7 shows that the EQ scales are the most stable across the statistics measured for the data in this study. This shows solid evidence of both reliability and homogeneity.

A Brief Review of Logistic Regression

The current study breaks each data set into two subsets, the Target group with classification equal to 1, and the Control group with classification equal to 0. In other words, our classification variable is binary. Note that one may consider more than two classifications using the logistic regression approach.

Generally speaking, there are a multitude of excellent references on logistic regression. This paper follows the work in [11], but also relies on the work in [1]. Suppose we have a single response variable y taking values in $\{0, 1\}$ and a single, continuous explanatory variable x. The corresponding logistic regression model is of the form

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{1}$$

where the notation exp denotes the usual exponential function with base e. The function $\pi : D \mapsto [0, 1]$ where D is an appropriate domain dependent on the explanatory variable x and $[0, 1]$ is the usual unit interval in $\mathbb{R}$.

According to [11], there are two main reasons for choosing the logistic distribution in (1). First, π is an extremely flexible and easily used function, and second, π lends itself to meaningful (clinical) interpretation. To see the utility of the function π note the following transformation, called the logit transformation.

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right). \tag{2}$$

Note that with a little algebra, $g(x) = \beta_0 + \beta_1 x$. This is useful in that the logit transformation of the logistic regression equation results in a linear expression with many of the desirable properties of the usual linear regression model.

One important difference between linear and logistic regression is that the error, which expresses an observation's deviation from the conditional mean, is no longer assumed to be normally distributed. Again following [11], we may express the value of the outcome variable given x as $y = \pi(x) + \epsilon$.

In this formulation, ϵ may take on one of two possible values. If $y = 1$, then $\epsilon = 1 - \pi(x)$ with probability $\pi(x)$, and if $y = 0$ then $\epsilon = -\pi(x)$ with probability $1 - \pi(x)$. In summary, ϵ follows a binomial distribution with probability given by the conditional mean $\pi(x)$.

The importance of the preceding discussion is that we can now readily construct the likelihood function of the above mentioned binomial distribution. For values of $y = 1$ given x the contribution to the likelihood function is $\pi(x)$ and the contribution for values of $y = 0$ given x the contribution is $1 - \pi(x)$. Thus, for any observation x_i, the contribution to the likelihood function is given by

$$\pi(x_i)^{y_i} \left[1 - \pi(x_i)\right]^{1-y_i}. \tag{3}$$

Note that (3) reduces to $\pi(x_i)$ or $1 - \pi(x_i)$ depending on the value of y_i given the choice of

x_i. One assumption in logistic regression is that the observations are independent and hence the likelihood function is given by the product of the individual terms given in (3):

$$\ell(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i} \left[1 - \pi(x_i)\right]^{1-y_i} \tag{4}$$

There is one more step involved to obtain the desired result. In all parametric regression approaches, there is an underlying optimization. This usually entails some form of differentiation. In the case at hand, (4) now requires differentiation with respect to the parameters $\boldsymbol{\beta}$ and a solution of the resulting equations. However, differentiation of products of functions is quite difficult compared to differentiation of sums of functions. This leads to a heavy computational cost. Hence, it is advantageous to construct the log likelihood function by taking the logarithm of (4) and using the appropriate properties of the logarithmic functions, namely that $\ln(f \cdot g) = \ln(f) + \ln(g)$ and $\ln(f^g) = g\ln(f)$. Hence,

$$\begin{aligned} \mathcal{L}(\beta) = \sum_{i=1}^{n} \{ & y_i \ln(\pi(x_i)) \\ & + (1 - y_i)(\ln(1 - \pi(x_i)))\}. \end{aligned} \tag{5}$$

defines the log-likelihood function.

The problem at hand is now to optimize (5) with respect to the parameters $\boldsymbol{\beta}$. It should be noted that while the parameters $\boldsymbol{\beta}$ are not explicitly present in (5), one may substitute the definition of $\pi(x)$ from (1) into (5) to see that (5) is, in fact, a function of the parameters $\boldsymbol{\beta}$.

An extension of logistic regression that may be useful in classification problems is that of multinomial logistic regression. As a quick example, suppose that the response variable now may take on any of 3 possible values, $\{0, 1, 2\}$. In this case, one may define the conditional probabilities of each outcome category as follows:

$$P(y = 0|x) = \frac{1}{1 + \exp(g_1(x)) + \exp(g_2(x))}, \tag{6}$$

$$P(y = 1|x) = \frac{\exp(g_1(x))}{1 + \exp(g_1(x)) + \exp(g_2(x))}, \tag{7}$$

and

$$P(y = 2|x) = \frac{\exp(g_2(x))}{1 + \exp(g_1(x)) + \exp(g_2(x))}, \tag{8}$$

where

$$g_i(x) = \beta_{i0} + \beta_{i1}x_1 + \ldots \beta_{in}x_n. \tag{9}$$

In (9) the index i runs from 1 to the number of categories present (2 in this example), and n represents the number of independent variables present.

There is a similar derivation of the log likelihood function to that in (5) and a maximum likelihood estimation process is used to find the coefficients (β_{ij}).

The utility of the multinomial logistic regression technique is for a case similar to predicting the likelihood of a student with a given set of characteristics to pass a given course with a particular grade level. This process could also be useful in constructing a predictive model that would rank a group of sales employees into two categories, one category representing high performers and the other category representing low to average performers. The third category may be a random sample of the general population for differentiation purposes.

Confusion Matrices, ROC, and AUC

This paper relies on the concepts of the receiver operating characteristic (ROC) curve and various derived statistics that are based on the information contained in the confusion matrix. For a more detailed version of the content of this section, one may consult [7]. For some critiques of the approach with additional options offered one may be interested in [15].

The confusion matrix is an organizational tool that contains information on the correct and incorrect classification counts of the model of interest. The goal here is to predict membership in a group. We may classify a prediction based on whether it correctly identifies membership. In this way, we wish to know the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

A TP is assigned to a prediction if the prediction assigns membership to the group of interest and the individual belongs to the group of interest. FP is assigned to a prediction that assigns membership to the group of interest and the individual does not belong to the group of interest. Similarly, TN is assigned to a prediction if the prediction assigns membership to the alternate to the group of interest and the individual does not belong to the group of interest. Finally, FN is assigned to a prediction if the prediction assigns membership to the alternate to the group of interest and the individual belongs to the group of interest.

We organize this information in the confusion matrix as follows. Let **p** and **n** denote the predicted outcomes and let **Y** and **N** denote actual membership. Further, let P' and N' denote the total positive and negative predictions, respectively. In other words, P' is the sum of TP and FP while N' is the sum of FN and TN. Figure 1 displays the resulting information contained in a confusion matrix. Note that $Y = P'$ and $N = N'$, by definition. In other words, the sum of TP and FN is the actual number of members in the group of interest and similarly for the sum of FP and TN.

		Prediction outcome	
		p	**n**
actual value	**Y**	True Positive	False Positive
	N	False Negative	True Negative
	total	P'	N'

Figure 1: Confusion Matrix

What has not been specified is how one determines whether a prediction correctly identifies membership. To address this, one must first define at what level one considers a prediction to be correctly classifying. For example, suppose we have 100 individuals and 25 of them belong to a group of interest. A reasonable cutoff would be to state that if the model assigns a probability of belonging to the group of interest greater than or equal to 0.25 we would consider that to be a positive. If the individual with a probability of at least 0.25 actually belongs to the group of interest, we increment the TP counter by one, otherwise we increment the FP counter by one. Similarly for probability less than 0.25 and the negative counters.

It is also reasonable to consider the cutoff of 0.50, especially if one knows that the sample population does not accurately depict reality. The fact is that it is highly unlikely for the modeler to know what the true population breakdown is. Therefore, we consider many different cutoff points which leads to the concept of the ROC curve. We first need to define some of the essential statistics, some of which are directly used in this paper and others are for information only.

First we define the true positive rate and the false positive rate. The true positive rate is the ratio of correctly identified positives to the total number of actual positives. In terms of the confusion matrix we have

$$TPR = \frac{TP}{P'}. \tag{10}$$

Similarly, the false positive rate is the total number of false positives divided by the total number of negatives.

$$FPR = \frac{FP}{N'}. \tag{11}$$

Some quantities of interest may be accuracy, the total correct classifications divided by the sample size:

$$Acc = \frac{TP + TN}{P' + N'}, \tag{12}$$

precision, the true positives divided by the total number of positives, true and false:

$$Pre = \frac{TP}{TP + FP}, \tag{13}$$

and a measure of overall fit called the F-measure

$$F = \frac{2}{\frac{1}{Pre} + \frac{1}{Acc}}. \tag{14}$$

Going back to the example mentioned before, we have a group of 100 total and 25 we are interested in identifying using some method. Suppose we have the following confusion matrix:

Table 8: Example Confusion Matrix

17	12
8	63

We may compute the above mentioned statistics.

$$TPR = \frac{17}{25} = 0.68, \tag{15}$$

$$FPR = \frac{12}{75} = 0.16, \tag{16}$$

$$Acc = \frac{17 + 63}{100} = 0.80, \tag{17}$$

$$Pre = \frac{17}{17 + 12} = 0.59, \tag{18}$$

and

$$F = \frac{2}{\frac{1}{.59} + \frac{1}{.8}} = .68. \tag{19}$$

We note that the above example is for a single paring of TPR with FPR. Each of these rates has been computed for a single classification point, the examples I gave were 0.25 and 0.50 for the fictitious data in Table 8. The previous quantities are not based on a real classification model.

To move to the concept of the ROC curve, we consider not just a single point, but all possible values we could use as a judgement of how well the model performs at each possible value in the interval $[0, 1]$. Since there are infinitely many possibilities, we take a reasonably sized discrete sample and compute a confusion matrix for each value we choose.

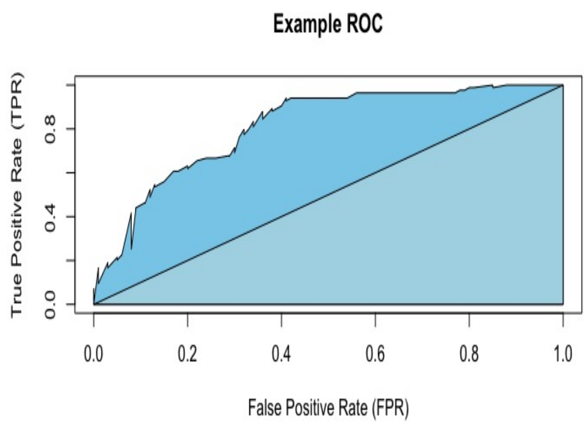

Figure 2: Example ROC Curve

For each confusion matrix we may compute any or all of the aforementioned quantities. However, we are particularly interested in the relationship between the true positive rate and the false positive rate. We consider how the functional relationship behaves, TPR as a function of FPR by considering these as ordered pairs of data (FPR_i, TPR_i), $i = 1, \ldots, n$, in the unit square in $\mathbb{R}^2$.

Visualization of this data is done by plotting the ordered pairs of data with the FPR_i values along the x-axis and the TPR_i values on the y-axis. This plot is known as the ROC curve. Figure 2 shows an example plot. The line separating the two shades of blue is the line $y = x$ and represents the ROC curve of a random selection process. In other words, if we randomly classified individuals the line $y = x$ represents the relationship between the FPR and TPR values.

Values of an ROC curve that lie above the line $y = x$ denote a classification algorithm that is superior to a random selection process with the reverse holding for points below the line. The example in Figure 2 shows a solidly performing algorithm. It is desirable to have a quantification of how much better the algorithm performs than the random approach. This value is given to us by the area under the curve (AUC), meaning the area under the classification ROC curve.

The line $y = x$ forms a right triangle with the x-axis and the vertical line $x = 1$. The area under this curve on the interval $[0, 1]$ is thus 0.50. If we can show the area under the ROC curve of interest is greater than or equal to 0.50, we have shown that the classification algorithm in question performs better than the random selection approach.

The solution to this problem is to integrate the function defining the ROC curve. In other words,

$$AUC = \int_0^1 f(x)dx \tag{20}$$

where $f(x)$ is a functional representation of the FPR-TPR relationship.

We do not have an explicit form for $f(x)$ generally speaking, but we do have techniques to integrate discrete functions that we may apply. For one such technique, the interested reader may consult Algorithm 1 in [7].

A Brief Introduction to Principal Component Analysis

This section provides a brief review of principal component analysis and is based mostly on the second chapter of [6].

Generally speaking, both the Style Insights® and Motivation Insights® portions of the Talent Insights® assessment are forced choice assessment questionnaires. The following fact is true of all forced choice assessments. If we set r to be the number of items in a forced choice block on an assessment then the off-diagonal elements of the correlation matrix of the scales of the forced choice assessment are given by

$$cor(X, Y) = -\frac{1}{r} \tag{21}$$

where X and Y represent any two scales measured by the assessment.

In a given modeling exercise it is desirable to use explanatory variables that are correlated with the response variable but not with other explanatory variables in as much as is possible to avoid over fitting. Given that the assessment scales are naturally correlated by the nature of the assessment, it may be possible to retain the information provided by the explanatory variables but do so in a way that removes as much correlation as possible.

One way of approaching this problem is to consider principal component analysis (PCA). PCA is a variable reduction process that uses linear combinations of the original variables to create a, presumably smaller, set of variables that accounts for as much of the variation in the original variables as possible.

Mathematically, if we let X_i, $i = 1, \ldots, n$ denote the set of original variables we define

$$PC_1 = \sum_{i=1}^{n} w_{(1)i} X_i \tag{22}$$

where the weights $w_{(1)i}$ are chosen to maximize the ratio of the variance of PC_1 to the total variation subject to the following constraint

$$\sum_{i=1}^{n} w_{(1)i} = 1. \tag{23}$$

We then compute the second new vector PC_2 which is now orthogonal to PC_1 which accounts for the maximum amount of the remaining total variation not accounted for by PC_1. This process continues until we have computed n new vectors that form an orthogonal basis for the space spanned by the original set of n explanatory variables. At this point, the art in the phrase "art and science" comes into play.

We choose the first m new variables $PC_1, \ldots, PC_m$ such that as much variation as possible is accounted for in the following sense. When adding the next variable to the set to be considered no longer adds a significant amount of variation accounted for, that next variable is not adding to the overall information to be considered. Exactly what the cutoff should be is a matter of some trial and error to see what works best in a given situation.

One positive takeaway from the PCA approach is that we now have a set of orthogonal variables and the probability of overfitting the model is much smaller. A negative of this approach is that the interpretation of the new variables is not necessarily straightforward.

Job Description Classification Results

This section presents some of the results of the current study. We begin with an application of the logit transformation presented in (2) as a tool to help determine whether variables of interest are candidates for inclusion in a logistic regression model.

Figure 3 presents a good example of a linear fit of the log odds of membership in an interval of a score on a scale, a desired property for inclusion. This particular case shows the Natural Dominance scale from the CEO data set from the Style Insights assessment. The log odds are computed using (2) where $\pi(x)$ is replaced by p and p represents the conditional probability of having a score in the interval of interest given membership in the group of interest.

The data shown in Figure 4, in contrast to the results of the Natural Dominance graph in Figure 3, show no linear relationship whatsoever. This data represents the Intentional scale from the Motivation Insights portion of the CEO data. The takeaway is that the Natural Dominance scale is a reasonable candidate for direct inclusion in a logistic regression model for the CEO data while the Intentional data appears to not be a candidate.

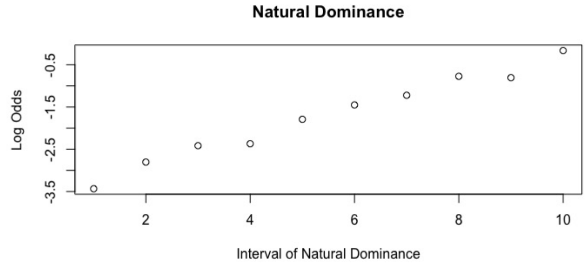

Figure 3: Natural Dominance Log Odds Plot

The linear relationship of the log odds of the data to membership in a group and bucket combination determines both the relative potential strength of the variable as a predictor and the directionality expected for any coefficient related to that variable in any logistic regression approach. As an example, one would expect that the Natural Dominance variable has a positive coefficient given its strong positive linear relationship. A failure to maintain this directionally correct relationship in the coefficients could be cause for removal of a variable from consideration.

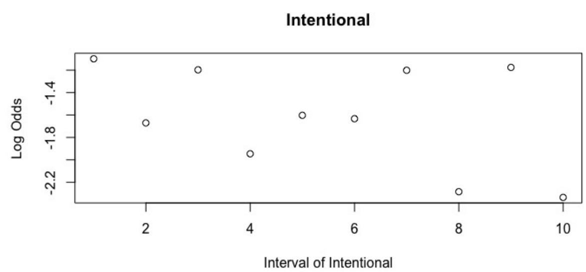

Figure 4: Intentional Log Odds Plot

As noted in a previous section, the ROC analysis is concerned with determining the relationship between the true positive rate (TPR) and the false positive rate (FPR). Further, we wish to determine whether the information taken from the assessment variables generates identification rates better than a random sampling technique.

This may be done in at least two ways. The first is to visually inspect the ROC curve and its relationship to the line $y = x$ in the plane. The second is to compute the area under the ROC curve and compare that to the area under the curve generated by the line $y = x$ on the interval $[0, 1]$. Recall that this area is equal to 0.50.

Examples of both methods are presented in a previous section of this paper. Additionally, many statistics are available to be compared across data sets and to help determine additional information about the quality of the fit. Given that this paper is interested in determining whether relationships exist between variables of interest, as opposed to attempting to generate a fully predictive model and testing its accuracy, the authors choose to present the Brier statistic, see, e.g., [15].

First presented is the area under the curve (AUC) statistic. Table 9 shows the results of the analysis applied to the CEO data set. The results are fairly consistent across the data subsets with Behaviors and All DISC (Behaviors with Behavioral Hierarchy included) scoring nearly identically, and showing a slight advantage over the scoring of the Motivators data subset. When considering the full TriMetrixEQ data set, we see an increase in AUC of 14.5% (0.71/0.62-1), from a low of 0.62 for Motivators and EQ alone to a high of 0.71 with all variables included.

Table 9: CEO Sample
Area Under Curve

Subset Considered	No PCA	PCA
Behaviors	0.65	0.65
All DISC	0.65	0.65
Motivators	0.62	0.62
DISC + Motivators	0.67	0.67
EQ	0.62	0.62
All	0.71	0.71

The Brier statistics show very little in the way of differences between the models, with only a slight change in the all variables version.

Figure 5 presents the ROC curve compared to the random selection technique. The curve is smooth throughout and clearly lies above the line $y = x$ consistent with the values of AUC being consistently above 0.50 as shown in Table 9.

Table 10: CEO Sample
Brier Statistic

Subset Considered	No PCA	PCA
Behaviors	0.19	0.19
All DISC	0.19	0.19
Motivators	0.20	0.19
DISC + Motivators	0.19	0.20
EQ	0.20	0.20
All	0.18	0.18

The authors would like to draw a comparison, again, to an earlier, similar study conducted on Serial Entrepreneurs. In this case, we draw the comparison between % Correctly Classified as presented in [8] and the AUC values in Table 9. This comparison is presented in Table 11. Note that the column Correct ID is a real number version of the % Correctly Classified as reported in the original work.

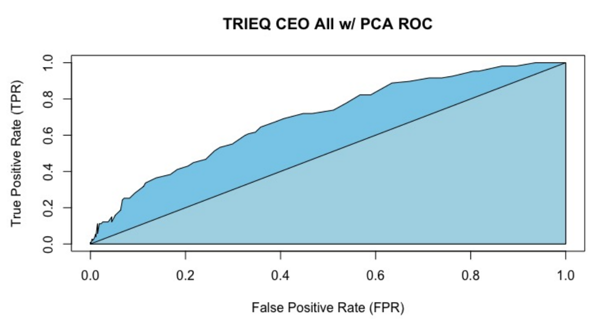

Figure 5: CEO ROC Curve: All Variables

There are a few things to make note of. First, there is no generalization between identification problems based solely on the variables present. In fact, [8] considered data from a TriMetrix Legacy DNA® assessment, collected in 2010, and this study which considered the TriMetrixEQ assessment for which data was collected between April and September, 2018. Similarly, this work does not differentiate between Adapted and Natural Behaviors while [8] considers them separately. The AUC scores presented in Table 11 are based on the PCA column of Table 9. Finally, the earlier study used a set of volunteer, self-described serial entrepreneurs while the current study used a selection of self-report Chief Executives as collected in a demographics bundle.

Table 11: CEO v Entrepreneur Comparison

Subset	AUC	Correct ID
All DISC	0.65	0.71
Motivators	0.62	0.67
All Variables	0.71	0.75

What the authors do find interesting are the similarities between the performance of a very similar modeling approach (logistic regression) on two separate and assumed distinct data sets. To elaborate, both groups of individuals are self identified as Serial Entrepreneurs or CEO. Serial Entrepreneur is defined as one who has started more than one business with no reference to the success or failure of any business started by the individual. CEO is an abbreviation of a job description in the O*Net classification and is used to denote Chief Executive.

While one may expect some similarities to exist, there is by no means any requirement for any specific qualities to be adhered to by any individual starting a business or by anyone advancing through the ranks of a company to attain a C-level position.

When comparing the results presented in Table 11 it is interesting to see that the different subsets considered perform so similarly. Additionally, the hierarchy is generally respected in that each individual subcomponent of TriMetrixEQ (Behavior and Motivator data) performs approximately the same across the subsets and across the two different studies, and combining the data together provides a better result. This final piece of information is critical.

A conclusion of [8] is that additional assessments provide additional information. The results of the current study appear to support that conclusion and support continued evidence of the relationship between TriMetrixEQ variables and external variables, in this case Chief Executives as self-reported and based on the O*Net job classification model.

As a final note, one may question the comparison of the AUC to the % Correctly Identified. To address this, note the the % Correctly Classified, as presented in [8], is based on the Confusion Matrix as defined earlier in this paper. For any specific value $x \in [0, 1]$ and a probability model, one may define a Confusion Matrix. For the results in [8], the assumption was made that the ratio of membership to non-membership in the set of interest was known and used to generate the Confusion Matrix (50%).

The ROC analysis presented in the current study are based on using a discrete approximation of the ROC curve, which is the functional relationship between the true positive rate and the false positve rate. The AUC statistic is based on a discrete approximation of the ROC curve and generates and approximation of the area under the true ROC curve. To be explicit, denote by $f : \mathbb{R} \mapsto \mathbb{R}$ any continuous function defined on the real line. Then,

$$\frac{1}{b-a} \int_a^b f(x)dx \tag{24}$$

defines the average value of f over the interval $[a, b] \subseteq \mathbb{R}$. If we let f be a function representing an ROC curve, we then have

$$AUC = \frac{1}{1-0} \int_0^1 f(x)dx = \int_0^1 f(x)dx. \tag{25}$$

Given that the % Correctly Identified from [8] is simply a point on the ROC curve, we can conclude that the AUC values discussed in this paper and the classification percentages from [8] may be compared as presented above.

Summary and Future Work

This paper has shown that the use of a logistic regression algorithm as an identification technique successfully outperforms the standard random selection technique it is measured against in a statistically significant way. This study is limited to a small population of self-described CEOs. While the data set is smaller than the researchers would like, the results are consistent with other studies conducted by TTI Success Insights.

With the inclusion of demographic information as a standard option in for all assessments, gathering information and data will continue for the foreseeable future. This snapshot is based on the CEOs from the TriMetrixEQ assessment and contains only 107 total respondents. A quick look at the Talent Insights® assessment data for the same period shows a population of 714 self-reported CEOs. According to the O*NET website, there are over 1,000 occupations listed. Continued studies establishing the relationships between the TTI Success Insights assessment variables and any external variables for which data is available.

References

[1] Alan Agresti. *Categorical Data Analysis, Third Edition*. Wiley, New York, 2012.

[2] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). Standards for educational and psychological testing, 2014.

[3] Lee Anna Clark and David Watson. Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3):309–319, 1995.

[4] Lee J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334, 1951.

[5] Robert F. DeVellis. *Scale Development: Theory and Applications*. Sage, Thousand Oaks, CA, 4th edition, 2017.

[6] William R. Dillon and Matthew Goldstein. *Multivariate Analysis: Methods and Applications*. Wiley Series in Probability and Mathematical Statistics, 1984.

[7] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

[8] Eric T. Gehrig. Classification of Serial Entrepreneurs via Logistic Regression: A Case Study. White Paper, October 2017.

[9] Kurt F. Geisinger, Bruce A. Bracken, Janet F. Carlson, Jo-Ida C. Hansen, Nathan R. Kuncel, Steven P. Reise, and Michael C. Rodrigues, editors. *APA Handbook of Testing and Assessment in Psychology Volume 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology*, 750 First Street NE, Washington, D.C. 20002-4242, 2013. American Psychological Association.

[10] Louis Guttman. A basis for analyzing test-retest reliability. *Psychometrika*, 42(4):252–282, 1945.

[11] David W. Hosmer and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley and Sons, Inc., 2000.

[12] C. Hoyt. Test reliability estimated by analysis of variance. *Psychometrika*, 6:153–160, 1941.

[13] G. F. Kuder and M. W. Richardson. The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160, 1937.

[14] Randall D. Penfield. Item Analysis. In Kurt F. Geisinger, Bruce A. Bracken, Janet F. Carlson, Jo-Ida C. Hansen,

Nathan R. Kuncel, Steven P. Reise, and Michael C. Rodrigues, editors, *APA Handbook of Testing and Assessment in Psychology Volume 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology*, chapter 7, pages 121–138. American Psychological Association, 750 First Street NE, Washington, D.C. 20002-4242, 2013.

[15] David M. W. Powers. Evaluation: From Precision Recall, and F-Factor to ROC, Informedness, Markedness, and Correlation. Technical report, Flinders University of South Australia, School of Informatics and Engineering, December 2007.

[16] Stephan G. Sireci and Tia Sukin. Test Validity. In Kurt F. Geisinger, Bruce A. Bracken, Janet F. Carlson, Jo-Ida C. Hansen, Nathan R. Kuncel, Steven P. Reise, and Michael C. Rodrigues, editors, *APA Handbook of Testing and Assessment in Psychology Volume 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology*, chapter 4, pages 61–84. American Psychological Association, 750 First Street NE, Washington, D.C. 20002-4242, 2013.