

2017

## Communicating Criterion-Related Validity Using Expectancy Charts: A New Approach

Jeffrey M. Cucina

*U.S. Customs and Border Protection*, [jcucina@gmail.com](mailto:jcucina@gmail.com)


Julia L. Berger

*Bowling Green State University, U.S. Customs and Border Protection, & ProMedica*, [jlberger@bgsu.edu](mailto:jlberger@bgsu.edu)

Henry H. Busciglio

*U.S. Customs and Border Protection*, [hhbusciglio@aol.com](mailto:hhbusciglio@aol.com)

Follow this and additional works at: <https://scholarworks.bgsu.edu/pad>

 Part of the [Industrial and Organizational Psychology Commons](#), and the [Quantitative Psychology Commons](#)

### Recommended Citation

Cucina, Jeffrey M.; Berger, Julia L.; and Busciglio, Henry H. (2017) "Communicating Criterion-Related Validity Using Expectancy Charts: A New Approach," *Personnel Assessment and Decisions*: Vol. 3 : Iss. 1 , Article 1.

DOI: 10.25035/pad.2017.001

Available at: <https://scholarworks.bgsu.edu/pad/vol3/iss1/1>

This Research Article is brought to you for free and open access by the Journals at ScholarWorks@BGSU. It has been accepted for inclusion in Personnel Assessment and Decisions by an authorized editor of ScholarWorks@BGSU.

# COMMUNICATING CRITERION-RELATED VALIDITY USING EXPECTANCY CHARTS: A NEW APPROACH

Jeffrey M. Cucina<sup>1</sup>, Julia L. Berger<sup>2</sup>, and Henry H. Busciglio<sup>1</sup>

1. U.S. Customs and Border Protection

2. U.S. Customs and Border Protection, Bowling Green State University

## ABSTRACT

### KEYWORDS

Expectancy Charts,  
Validity,  
Bivariate Normal  
Distribution

Often, personnel selection practitioners present the results of their criterion-related validity studies to their senior leaders and other stakeholders when trying to either implement a new test or validate an existing test. It is sometimes challenging to present complex, statistical results to nonstatistical audiences in a way that enables intuitive decision making. Therefore, practitioners often turn to expectancy charts to depict criterion-related validity. There are two main approaches for constructing expectancy charts (i.e., use of Taylor-Russell tables or splitting a raw dataset), both of which have considerable limitations. We propose a new approach for creating expectancy charts based on the bivariate-normal distribution. The new method overcomes the limitations inherent in the other two methods and offers a statistically sound and user-friendly approach for constructing expectancy charts.

Imagine that you are a practitioner and have just conducted a criterion-related validity study on a new personnel selection test. Your organization's top leadership asks for a quick summary of the findings. How would you convey the predictive ability of the test on a single slide in a presentation without having to delve into statistical terminology? In the academic literature, criterion-related validity coefficients are often used as a measure of predictive ability of a test. However, practitioners may find it challenging to translate the meaning of a criterion-related validity coefficient to a nontechnically-savvy audience in a manner that would allow the audience to quickly make informed decisions. We have found expectancy charts to be useful for these purposes. These charts can convey the predictive ability of a test in a single slide with little explanation required.

To help practitioners translate the results of their criterion-related validity studies, we set out to meet the following goals: (a) to describe how expectancy charts can assist personnel selection practitioners in translating complex, technical concepts such as criterion-related validity to nontechnically-savvy audiences and (b) present a new methodology for constructing expectancy charts. We also provide R syntax that practitioners can use to compute more accurate expectancy values. We begin our paper by reviewing existing approaches to creating expectancy charts and highlighting some potential issues with these approaches. We then present our new approach and explain how to

implement it in a step-by-step format. We also discuss the inherent assumptions and limitations of our approach and cover some special situations in which it may be used (e.g., multiple-hurdle selection systems).

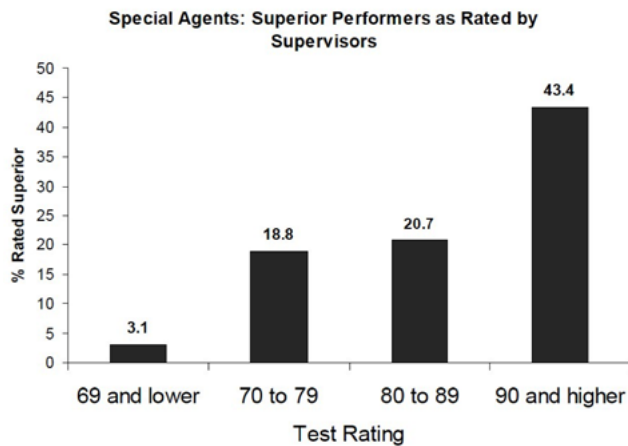
### Expectancy Charts

Expectancy charts are bar charts that illustrate the relationship between a range of predictor scores, such as personality, and a range of criterion scores, such as job performance. [Figure 1](#) illustrates an expectancy chart for the relationship between ranges of test scores and the percentage of special agents who were rated as superior by their supervisors. For example, of those special agents with test scores of 90 or higher, 43.4% were rated as superior, compared to 3.1% for those with test scores of 69 or lower.

Julia L. Berger is now at Aptima, Inc and Henry H. Busciglio is now retired. The views expressed in this paper are those of the authors and do not necessarily reflect the views of U.S. Customs and Border Protection or the U.S. Federal Government. The authors would like to thank Philip T. Walmsley, Kimberly J. Wilson, and Chihwei Su for their valuable comments and suggestions on this article. A portion of this article was presented at the 2016 annual meeting of the Society for Industrial and Organizational Psychology (Division 14 of the American Psychological Association).

### Corresponding author:

Jeffrey M. Cucina  
Address: 1400 L Street, NW 7S39  
Washington, DC 20229-1145  
Email: jcucina@gmail.com  
Phone: 202-863-6298



**FIGURE 1.**

This expectancy chart shows that special agents who have scored 90 and higher on a test are more likely to be rated superior by their supervisors. Note that this expectancy chart was computed using raw data and was reproduced from [Simpson, Nester, and Palmer \(2007\)](#).

**Nontechnical Explanation:** This test predicts who will be a superior performer on the job. Of those employees with a test score of 69 or lower, only 3.1% were rated as superior. In contrast, 43.4% of the employees with a test score of 90 or higher were rated as superior. To identify superior performers, supervisors confidentially rated the performance of their employees. Those employees who were in the top 25% were identified as superior.

Thus, expectancy charts allow practitioners to present complex statistical relationships in an easy fashion. As research has long shown, visualizations aid memory and comprehension because they help build mental models, whereas text and numbers do not ([Glenberg & Langston, 1992](#)).

### Taylor-Russell Table Approach

When deciding to construct an expectancy chart, a practitioner may discover two dominant approaches in the literature. The first approach uses the Taylor-Russell tables to obtain expectancies ([Taylor & Russell, 1939](#)). This requires three inputs: the validity coefficient, the proportion of applicants who will be selected top-down based on their test scores, and the base rate proportion of current employees who are satisfactory performers. With these three values, one can use the Taylor-Russell tables to determine the proportion of individuals who will be satisfactory performers if the test is used. For example, if the correlation between a test and job performance is .70, the proportion of satisfactory performers is .50, and the top 20% of applicants will be selected based on their test scores, then the expected proportion of satisfactory performers will increase from .50 to .90 when the test is used. There are some limitations to

the Taylor-Russell table approach. These tables only provide expectancies for ranges of test scores that go from a specified test score to the maximum possible score on the test (e.g., 70 or above, 80 or above). Thus, these tables do not provide expectancies for other types of test score ranges (e.g., 70 to 79, 80 to 89). Additionally, the three inputs must be rounded to increments of .05 or .10 when using the tables.

### Raw Data Approach

The second approach uses a raw dataset to compute expectancies. In this approach, the predictor and criterion are recoded into groups (e.g., four equally sized groups, or quartiles) or into specific ranges of scores (e.g., 70-79). Next, cross-tabulations are run on the data to obtain the expectancy values. For example, in [Figure 1](#), the raw data for the predictor were divided into four groups (i.e., 69 and lower, 70-79, 80-89, and 90 and higher) and the percentage of employees in each group who were rated superior was recorded. There are two limitations to this approach. First, it cannot incorporate corrections for criterion unreliability and range restriction. It is well recognized in the personnel selection literature that the relationship between an assessment test and performance in a raw dataset is artificially lowered due to these effects ([Nunnally & Bernstein, 1994](#); [Schmidt & Hunter, 1996, 1998](#)). Thus, expectancy charts computed using raw data underestimate the relationship between two variables.

Second, unless the sample size of the raw data is very large, there may not be enough cases to accurately compute expectancy values when the data are split into multiple groups. The sample sizes for criterion-related validity studies are often guided by a power analysis for detecting a significant correlation not for splitting the data into different groups. To illustrate the inherent noisiness associated with using raw data to create expectancy charts, we conducted a Monte Carlo simulation. The details and full results on the simulation are provided in the Supplemental Materials. We simulated data for an observed validity of .26, which is the observed meta-analytic value for cognitive ability tests ([Hunter, Schmidt, & Le, 2006](#)). We generated 10,000 samples of 151 cases each. With 151 cases, there is a 90% power of detecting a significant correlation. Next, we generated expectancies using the raw datasets. To compute the expectancies, we divided the predictor into five equally sized groups and recorded the percentage of superior performers. We defined superior performers as those within the top 20% on the criterion.

As shown in the Supplemental Materials, the results suggest that there is considerable variability and inaccuracy in expectancy values using raw data. This is due to the sampling error associated with splitting the dataset of 151 cases into multiple groups. The expectancy values varied considerably from sample to sample and the ranges of observed

values were quite large. In the population for our Monte Carlo simulation, the relationship is entirely monotonic, which means that increases in test scores are always associated with increases in job performance. However, in 87.4% of samples, the expectancies suggested that the predictor–criterion relationship was nonmonotonic, meaning that job performance did not always increase with test scores. Even in those samples where the observed validity (.26) turned out to be identical to the true population observed validity (.26), there was still considerable variation in the expectancies, and 88.2% of these samples suggested a nonmonotonic relationship.

A nonmonotonic relationship suggests that higher test scores are not associated with better performance. For example, in one sample that had a validity of .26, the expectancy values for the five equally sized groups of test scores were 16.7%, 9.1%, 26.5%, 20.7%, and 23.3%. This suggests that individuals whose test scores were in the lowest 20% did better than those who were in the next 20%. Further, individuals whose test scores were in the middle 20% outperformed those in the other five groups. Looking at these expectancy charts, a decision maker might assume that the organization should focus hiring on individuals whose test scores were in middle category or that the test is not a consistent predictor of performance. In contrast, the values that should have been obtained are 10.8%, 15.6%, 19.2%, 23.4%, and 31.2%. These values show a clear positive trend and do not paint the false picture that performance goes up and down as test scores increase.

Furthermore, as mentioned before, the raw data approach does not allow for corrections for range restriction and criterion unreliability. To corroborate this, we computed the average expectancy value for individuals who scored in the top 20% on both the predictor and the criterion from the Monte Carlo simulation and obtained a value of 30.2%. Next, using the new methodology that we will soon describe, we computed the expectancy value after making the corrections. The resulting expectancy was 46.0%, which is noticeably larger than the value obtained using the raw data. Thus, using raw data appears to be problematic.

### A New Approach to Computing Expectancies

To address the limitations inherent in the Taylor-Russell (1939) and raw data approaches, we developed syntax to compute more accurate expectancy values. This approach uses the bivariate normal distribution, which can be viewed as a three-dimensional bell-shaped graph showing the distribution and relationship between two variables. Information on the mathematical details of our new approach is provided in the Supplemental Materials. Essentially, our syntax computes expectancy values by selecting different sections under the bivariate normal distribution. It is equivalent to computing expectancies using a dataset of enormous size with no criterion unreliability or range restriction.

Our new methodology has several advantages and in many cases improves upon the Taylor-Russell (1939) and raw data approaches. In comparison to the raw data approach, our new methodology allows selection researchers to make corrections for range restriction and criterion unreliability while also reducing the effects of sampling error. Although our approach does not correct for the sampling error associated with estimating validity, it does address the sampling error that is observed in expectancy values when the validity coefficient is held constant. As mentioned above, even when the validity coefficient was .26 in the random samples, splitting the data into multiple categories resulted in expectancy values that varied widely. Because our new approach is equivalent to splitting a dataset of infinite size, it results in no noticeable sampling error when criterion-related validity is held constant. Second, our new approach eliminates the nonmonotonic relationships in expectancy charts that are due to the sampling error associated with splitting up the raw data. Third, unlike the Taylor-Russell (1939) approach, this approach allows for the computation of expectancies for different ranges of test scores. Fourth, it allows for more precision than the Taylor-Russell tables, which only provide expectancies for a subset of all possible input values.

Fifth, this approach allows nonstatistically minded audiences to visualize the relationships between organizational variables (e.g., cognitive ability, job performance, training performance), which aids comprehension and memory (Glenberg & Langston, 1992). We have found that these charts can quickly and easily convey the relationship between test scores and performance to stakeholders, especially when compared to correlation coefficients, regression equations, and scatterplots. As most nonstatisticians are familiar with percentages, only short descriptions of the expectancy chart are needed. An example of a nontechnical explanation is provided in Figure 1. Sixth, it is also possible to create expectancy charts using summary data from published studies. Table 1 displays expectancy charts for some of the most oft-researched relationships in industrial-organizational psychology. To make these charts, we obtained meta-analytic correlations from previous studies and computed expectancies for four equally sized groups (just as in the steps shown below).

### Step-by-Step Instructions for Constructing Expectancy Charts

The syntax for our new approach is provided in Table 2, along with annotations explaining each line of code. The syntax uses an algorithm created by Miwa and colleagues (Mi, Miwa, & Hothorn, 2009; Miwa, Hayter, & Kuriki, 2003) that is implemented in an R package developed by Genz and colleagues (Genz et al., 2008; Hothorn, Bretz, & Genz, 2001). To run the R syntax, researchers should perform the steps shown below. A copy of the output for the example values in the steps below is provided in Figure 2.



1. Diagnose the data for nonnormality, including skewness, kurtosis, and the shape of the distribution of scores. [Tabachnick and Fidell \(2001\)](#) provide an explanation of how to conduct and interpret these tests. If nonnormality is detected, a transformation should be performed to make the data normal. Possible transformations include the inverse, logarithmic, or [Blom \(1958\)](#) transformations in SPSS.
2. Examine the relationship between the predictor and criterion for nonlinearity. This can be done using the curve estimation feature in SPSS. If nonlinearity exists, conduct a regression analysis using standard statistical packages like SPSS, save the predicted scores, and use those scores as the predictor in further analyses.
3. Obtain the uncorrected correlation between the predictor and criterion. If needed, make corrections for range restriction and unreliability. Obtain the upper and lower bounds of the confidence intervals around the correlation, if desired.
4. Determine the cutoff scores for the different ranges of predictor and criterion scores, and convert these to a z-scale (i.e.,  $M = 0$ ,  $SD = 1$ ). For example, if you wish to divide the range of scores into quartiles (i.e., four equally-sized groups), the z-score cutoffs for the 4 ranges of scores would be:  $-\infty$  to  $-.67$  for the bottom 25%,  $-.67$  to  $0$  for the lower middle 25%,  $0$  to  $+.67$  for the upper middle 25%,  $+.67$  to  $\infty$  for the top 25%.
5. Download and install R on your computer and then launch the program. Alternatively, go to the website [www.r-fiddle.org](http://www.r-fiddle.org), which can run R syntax without installing R.
6. If you installed R on your computer, download the `mvtnorm` R package. (This step is not needed if you are using [www.r-fiddle.org](http://www.r-fiddle.org).)
7. Copy and paste the syntax shown in [Table 2](#) into R and press enter.
8. Type the following text into R: *Expectancyfunc (Validity, PredLowerCut, PredUpperCut, CritLowerCut, CritUpperCut)*.  
Substitute the validity coefficient from Step 2 for “Validity,” the lower and upper predictor cutoffs from Step 3 for “PredLowerCut” and “PredUpperCut,” respectively, and the lower and upper criterion cutoffs from Step 3 for “CritLowerCut” and “CritUpperCut,” respectively. Use the text “Inf” in place of an  $\infty$  symbol.  
For example, if the criterion-related validity is .545, the predictor scores of interest range from 1.0 to  $\infty$ , and the criterion scores of interest range from 0.67 to  $\infty$ , then the following text should be typed into R: *Expectancyfunc (0.545, 1, Inf, 0.67, Inf)*
9. Press <Enter> and the expectancy value will appear along with the joint probability and x probability values.

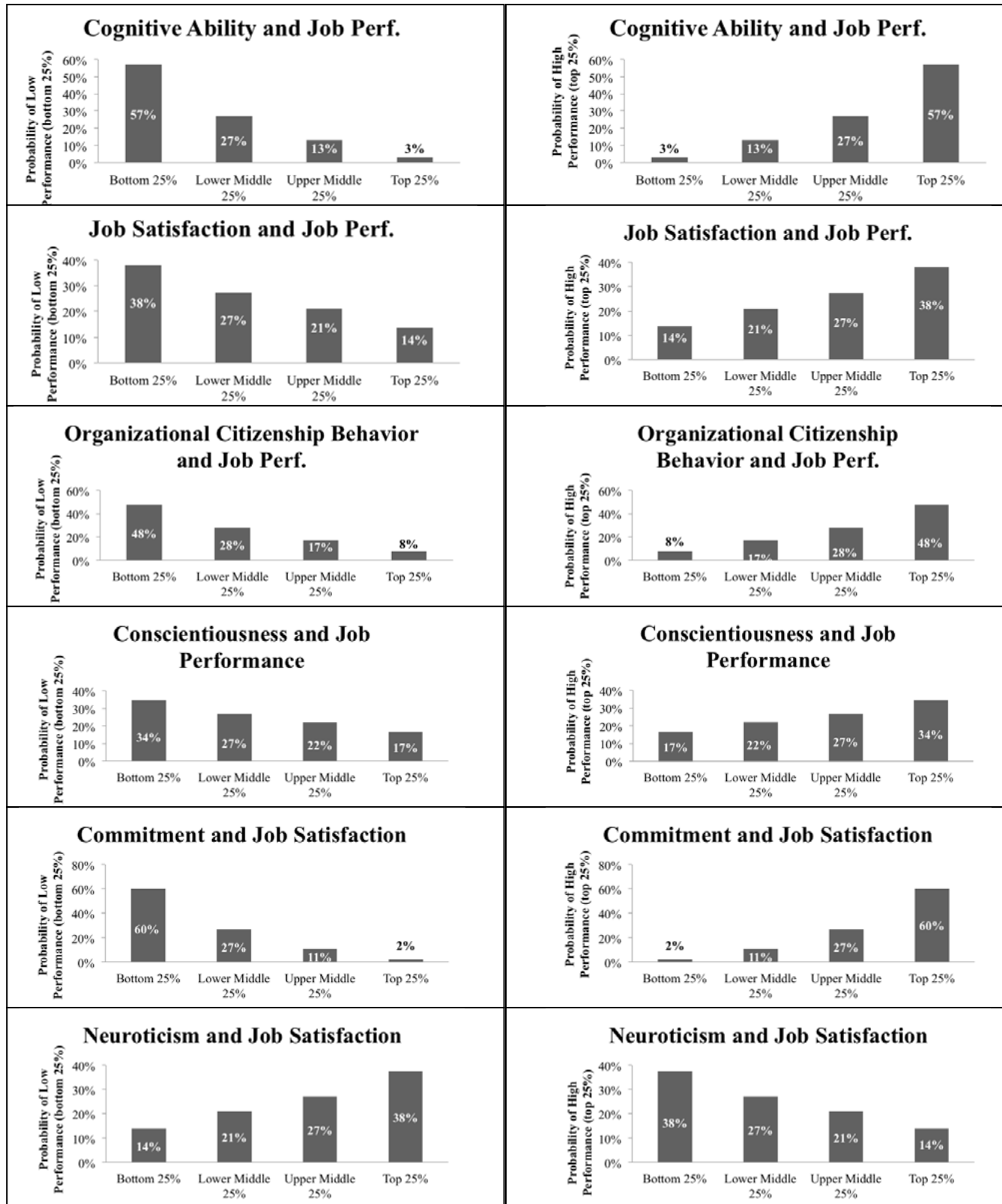
### Assumptions of the New Approach

We must mention that our new approach to creating expectancy charts makes a number of critical assumptions. The first assumption is that the two variables under study are normally distributed. There is some debate in the literature about the normality of variables ([Beck, Beatty, & Sackett, 2014](#)). Thus, it might be beneficial to check for nonnormality. Our syntax could then be applied and the score cutoffs could be transformed back to the nonnormal scale, if desired. If the data are nonnormal and no transformations are conducted, then the expectancy values will not be accurate. This is due to the fact that the algorithm behind our syntax computes the area under a bivariate normal distribution and not the area under a nonnormal distribution.

The second assumption is that the two variables must be linearly related. In general, most test scores tend to be linearly related to performance ([Coward & Sackett, 1990](#)), with the possible exceptions of personality predictors ([Carter et al., 2014](#)) and self-reported grades ([Arneson, Sackett, & Beatty, 2011](#)). Our syntax does not allow for the specification of the nonlinear terms, such as quadratic terms. However, it is possible to linearize the relationship by conducting a multiple regression analysis that includes curvilinear terms and then saving the predicted ( $\hat{y}$ ) values. The predicted values would have a linear relationship with the criterion and could be used as the predictor in our new approach. If a nonlinear relationship is present in the data but not considered when running our syntax, then the expectancy values will be incorrect. For example, if the relationship between the predictor and criterion begins to level off near the high end of the range of test scores, then the reported expectancy values for the high end of test scores will be higher than they should be.

Third, our approach assumes that the two variables are on an interval scale rather than an ordinal scale. Typically, individual items on a test or a criterion are on an ordinal scale and total scores are treated as an interval scale ([Nunnally & Bernstein, 1994](#)). Thus, if the predictor is a test score and the criterion is a sum total or average of multiple items, both variables can be treated as interval scales, and our approach can be used. However, if the predictor is an individual item such as a single personality item on a 5-point scale and/or the criterion is an individual item such as a single rating of job performance on a 5-point scale, then our new approach may not be valid. In our experience, most predictors and criteria used by practitioners are formed from multiple items, so this assumption may not be violated often in practice. Finally, our approach assumes that corrections for range restriction and criterion unreliability are appropriate. There indeed is some debate on this topic ([LeBreton, Scherer, & James, 2014](#); [Shen, Cucina, Walmsley, & Seltzer, 2014](#); [Viswesvaran, Ones, Schmidt, Le, & Oh, 2014](#)). We leave it to readers to decide where they stand on this debate and whether or not to make corrections and which values to apply when using corrections.

**TABLE 1.**  
*Expectancy Charts for Some of the Most Often Researched Relationships in Industrial-Organizational Psychology*



*Note.* The correlation coefficients used to create these expectancy charts were obtained from Paterson, Harms, and Credé (2012) and Schmidt, Shaffer, and Oh (2008).

TABLE 2.

*Computing the Volume Under the Bivariate-Normal Distribution*

| R script  | Annotation  |
|---|---|
| Expectancyfunc <- function (Validity, PredLowerCut, PredUpperCut, CritLowerCut, CritUpperCut) { | This creates a new function in R called Expectancyfunc. The function takes the criterion-related validity coefficient, the lower and upper cutoffs for the predictor score, and the lower and upper cutoffs for the criterion score as inputs. To represent positive or negative $\infty$ , “Inf” or “-Inf” can be used, respectively.  |
| library(mvtnorm)  | Before proceeding, the mvtnorm library must be downloaded and installed. This command line tells R that the mvtnorm library is being used.  |
| n <- 1000   | A dataset must be created before R can be run to conduct the analyses. This command tells R to create a dataset with 1,000 cases. The value n represents the number of cases and the symbol <- indicates that n should be set equal to 1,000.   |
| mean <- c(0, 0)   | In this line, the means for the two variables (which equal 0 when a standardized solution is used) are provided. Note that the values are presented parenthetically, separated by a comma, and preceded by the letter c. This syntax stores the means as a vector in R.   |
| lower <- c(PredLowerCut, CritLowerCut)  | This line assigns the lower z-score cutoffs for the predictor and the criterion to a vector.  |
| upper <- c(PredUpperCut, CritUpperCut)  | This line assigns the upper z-score cutoffs for the predictor and the criterion to a vector.  |
| corr <- diag(2)   | This line creates a 2-by-2 matrix with diagonal values of 1 and stores the matrix in the variable corr.   |
| corr[lower.tri(corr)] <- Validity<br>corr[upper.tri(corr)] <- Validity                          | In these two steps, the correlation between the two variables provided by the user is stored into the upper and lower triangles of the 2-by-2 correlation matrix.   |
| jtprob <- pmvnorm(lower, upper, mean, corr, algorithm = Miwa(steps = 128))                      | Here the pmvnorm command in the mvtnorm package is run; this is the command that is used for computing the volume under multivariate-normal distributions. As inputs, pmvnorm takes the upper and lower z-score cutoffs (which are vectors), the vector of means (which is set to 0), the correlation matrix, and the algorithm that is to be used. The algorithm statement specifies that the Miwa et al. (2003) algorithm should be used. The term “(steps = 128)” informs R that 128 grid points should be used. The output for this procedure is the joint probability between the predictor and the criterion – the volume under the bivariate-normal distribution between the lower and upper cutoffs. This probability is saved in the variable jtprob.  |
| jtprobOutput <- paste("Joint Probability: ", jtprob, sep="")                                    | This line creates a new string variable containing the value of jtprob along with a label. The term “sep=” indicates that there are no text separating the expectancy value and the % symbol.   |
| print(jtprobOutput)   | The previous steps saved the volume of the bivariate-normal distribution and added a label; this step prints that value, with the label, to the screen.   |
| <i>Computing the expectancy</i>   |   |
| xprob <- pnorm(PredUpperCut, mean=0, sd=1)-pnorm(PredLowerCut, mean=0, sd=1)                    | To compute the expectancy, we must obtain the proportion of cases that have a predictor value within the lower and upper cutoffs for the predictor. This is accomplished by computing the area under the univariate-normal distribution, which is the proportion of cases having predictor values within the upper and lower cutoffs. The pnorm command in R is used to compute this area and it takes the upper or lower predictor cutoff, mean (which is set to 0), and standard deviation (which is set to 1) as inputs. The proportion of cases that fall within the upper and lower cutoffs is obtained by subtracting the proportion of cases falling between the lower cutoff and $-\infty$ from the proportion of cases falling between the upper cutoff and $-\infty$ . This value is stored to a new variable, xprob. |
| xprobOutput <- paste("Predictor Probability: ", xprob, sep="")                                  | This line creates a new string variable containing the value of xprob along with a label.   |
| print(xprobOutput)  | This command prints the value xprob to the screen along with a label.   |
| expectancy <- paste(round(100*jtprob/xprob, 1), "%", sep="")                                    | The expectancy is computed by dividing the joint probability by the predictor probability. The expectancy is converted to a percentage using the syntax “100*.” In addition, this value is rounded to one decimal place, using the syntax “round(., 1).” Next, a percentage symbol is added using the “paste” command, which pastes the expectancy value and the % symbol (shown in the syntax using “%”) together into a string variable named “expectancy.”   |
| print(expectancy)   | This command prints the expectancy value to the screen.   |
| }   |   |

```

> Expectancyfunc <- function (validity, PredLowerCut, PredUpperCut, CritLC
wercut, CritUpperCut) {
+
+   library(mvtnorm)
+   n <- 1000
+   mean <- c(0, 0)
+   lower <- c(PredLowerCut, CritLowerCut)
+   upper <- c(PredUpperCut, CritUpperCut)
+
+   corr <- diag(2)
+   corr[lower.tri(corr)] <- validity
+   corr[upper.tri(corr)] <- validity
+
+   jtprob <- pmvnorm(lower, upper, mean, corr, algorithm = Miwa(steps =
128))
+   jtproboutput <- paste("Joint Probability: ", jtprob, sep="")
+   print(jtproboutput)
+
+   xprob <- pnorm(PredUpperCut, mean=0, sd=1)-pnorm(PredLowerCut, mean=
0, sd=1)
+   xproboutput <- paste("Predictor Probability: ", xprob, sep="")
+   print(xproboutput)
+   expectancy <- paste(round(100*jtprob/xprob, 1), "%", sep="")
+   print(expectancy)
+ }
>
> Expectancyfunc (0.545, 1, Inf, 0.67, Inf)
[1] "Joint Probability: 0.0906844166061603"
[1] "Predictor Probability: 0.158655253931457"
[1] "57.2%"

```

FIGURE 2.

This is the output (obtained using R Studio, 2015) from the R function shown in Table 2 with the syntax *Expectancyfunc* (0.545, 1, Inf, 0.67, Inf).

### Displaying Confidence Intervals and Corrections in Expectancy Charts

Statistically savvy audiences might inquire about the precision of the expectancy charts for a given study or about the impact of corrections for unreliability and range restriction on the expectancies. To address this, we propose a new format for expectancy charts that can be used to display confidence intervals (CIs) and corrections for unreliability and range restriction. Oftentimes, CIs are displayed on charts (showing mean differences) using error bars. The same approach can be applied to expectancy charts. The uncorrected and corrected correlation coefficients can also be portrayed using a traditional bar chart format. We depict both of these formats in Figure 3. Making corrections for range restriction and criterion unreliability impacts the expectancies. For example, in Figure 3, the uncorrected validity coefficient ( $\rho = .511$ ) indicates that of those individuals whose test score is in the top 25%, a total of 46% are superior performers in training (i.e., in the top 25%). After correcting for range restriction and criterion unreliability, the expectancy value increases to 52%, indicating that the original expectancy of 46% is actually an underestimate of the true expectancy. A format, such as that in Figure 3, can also be used to compare expectancies for the different predictors that an organization is considering for inclusion in a selection system.

### Using Expectancy Charts for Multiple-Predictor Selection Systems

It is common to think about expectancy charts as being applicable only to situations when there is a single predictor and a criterion. However, these charts can also be applied

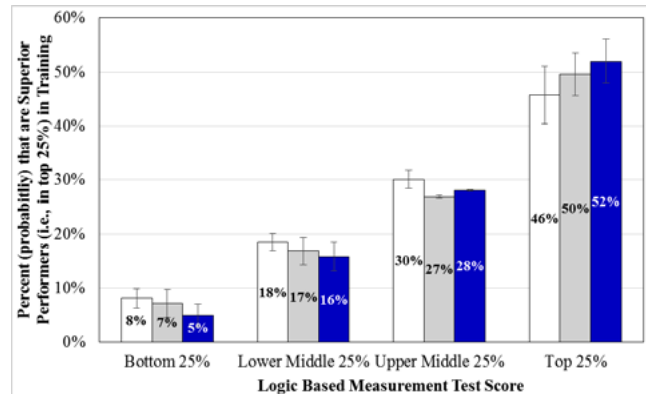


FIGURE 3.

New expectancy chart format showing confidence intervals as well as corrections for criterion unreliability and range restriction. The blue bar represents values corrected for criterion unreliability and range restriction, the gray bar represents values corrected for criterion unreliability, and the white bar represents the uncorrected values. This expectancy chart displays the expectancy percentages for the top 25% of individuals (using training performance as the criterion) and the quartiles of logic-based measurement (LBM) test scores. Note that the cutoff points for the quartiles of LBM scores are as follows:  $-\infty < X < -.67$ ;  $-.67 < X < 0$ ;  $0 < X < .67$ ;  $.67 < X < \infty$ . Similarly, the cutoff points for the criterion are  $.67 < X < \infty$ . Note that the z-scores of  $-.67$  and  $.67$  correspond to the 25th and 75th percentiles, respectively. The uncorrected validity coefficient ( $r = .457$ ), validity coefficient corrected for criterion unreliability ( $\rho = .511$ ), and validity coefficient corrected for criterion unreliability and range restriction ( $\rho = .600$ ) were drawn from Hayes, McElreath, and Reilly's (2002, 2003) meta-analysis. The upper and lower bounds of the 95% confidence intervals for the uncorrected correlation coefficient were corrected for unreliability and range restriction to obtain the corrected correlation's confidence intervals (see Schmidt & Hunter, 1996). The meta-analytic sample size ( $n = 6,711$ ) was too large to provide easily discernable confidence intervals in the chart; therefore, we arbitrarily set the sample size to 500 when computing the confidence intervals.

to each step in multihurdle personnel selection system provided that the range restriction corrections are made in an appropriate fashion. Consider an organization that has a two-step selection process, consisting of Test A in Step 1 and Test B in Step 2. In this situation, there are really two applicant pools. The first applicant pool consists of those applicants who participate in Step 1 and take Test A. The second applicant pool consists of those applicants who pass Step 1, participate in Step 2, and thus take Test B.

Suppose that the corrected validity for scores on Test A is .50. This is the estimate of what the validity of Test A would be for the first applicant pool if there was no range



restriction or criterion unreliability. Using our syntax, the expectancy values for the top 20% on the criterion and five equally sized groups of test scores are 4.2%, 10.2%, 16.7%, 25.4%, and 43.6%. Suppose that the corrected validity for Test B scores is .30. This estimate was computed for the second applicant pool, which are the applicants who passed Test A and participated in the second step. Using our syntax, the expectancy values are 9.6%, 14.8%, 19.0%, 23.8%, and 33.1%. Note that the validity of Test B is estimated only for those applicants who passed Test A, and it likely experiences incidental range restriction due to selection on Test A. If Test B were given in the first step, then its validity might be higher due to a lack of incidental range restriction. However, in this example, we are interested in the validity of Tests A and B within the steps in which these are given.

It is also possible to create expectancy charts for a composite score computed using two or more tests. For example, the meta-analytic multiple correlation between a composite of a general mental ability test and an integrity test with job performance is .65 (Schmidt & Hunter, 1998). Using our syntax, the expectancy values for this composite are 1.5%, 6.2%, 13.6%, 25.9%, and 53.0%. Thus, expectancy charts can be used for selection systems consisting of a single test or multiple tests.

## CONCLUSION

This paper presents expectancy charts as a useful way to display complex, statistical relationships to nontechnical-savvy audiences. It further presented a new methodology for creating expectancy charts and provided a step-by-step guide for implementing this methodology using R syntax. It is our hope that this article will make the creation of accurate expectancy charts easier for practitioners and researchers, and better facilitate the communication of information on the validity of assessments to stakeholders.

## REFERENCES

- Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the more-is-better and the good-enough hypotheses. *Psychological Science*, 22(10), 1336-1342.
- Beck, J. W., Beatty, A. S., & Sackett, P. R. (2014). On the distribution of job performance: The role of measurement characteristics in observed departures from normality. *Personnel Psychology*, 67(3), 531-566.
- Blom, G. (1958). *Statistical estimates and transformed beta variables*. New York, NY: John Wiley and Sons.
- Carter, N. T., Dalal, D. K., Boyce, A. S., O'Connell, M. S., Kung, M. C., & Delgado, K. M. (2014). Uncovering curvilinear relationships between conscientiousness and job performance: How theoretically appropriate measurement makes an empirical difference. *Journal of Applied Psychology*, 99(4), 564.
- Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology*, 75(3), 297.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2008). mvtnorm: Multivariate normal and t distributions. R package version 0.9-2. Available at <http://CRAN.R-project.org/package=mvtnorm>.
- Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, 31(2), 129-151.
- Hayes, T.L., McElreath, J.M., & Reilly, S.M. (2002, April). The criterion-related validity of logic-based tests of reasoning for personnel selection. In T. L. Hayes (Chair), *The validity of logic-based measurement for selection and promotion decisions*. Symposium presented at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.
- Hayes, T.L., McElreath, J.M., & Reilly, S.M. (2003). *The criterion-related validity of logic-based measurement and reasoning tests in public sector merit-based selection systems*. (Technical Report No. 03-01). Washington, DC: U.S. Customs and Border Protection.
- Hothorn, T., Bretz, F., & Genz, A. (2001). On multivariate t and Gauß probabilities in R. *R News*, 1(2), 27-29.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594.
- LeBreton, J. M., Scherer, K. T., & James, L. R. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology*, 7(4), 478-500.
- Mi, X., Miwa, T., & Hothorn, T. (2009). mvtnorm: New numerical algorithm for multivariate normal probabilities. *The R Journal*, 1(1), 37-39.
- Miwa, T., Hayter, A.J., & Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 223-234.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Paterson, T., Harms, P., & Credé, M. (2012, April). *The meta of all metas: 30 years of meta-analysis reviewed*. Poster presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- R Studio. (2015). *RStudio Desktop 0.98.1102*. Boston, MA: Author.
- Schmidt, F. L., & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199-223.
- Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L., Shaffer, J. A., & Oh, I. S. (2008). Increased accuracy for range restriction corrections: Implications

- for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61(4), 827-868.
- Shen, W., Cucina, J. M., Walmsley, P. T., & Seltzer, B. K. (2014). When correcting for unreliability of job performance ratings, the best estimate is still. 52. *Industrial and Organizational Psychology*, 7(4), 519-524.
- Simpson, R.W., Nester, M.A., & Palmer, E. (2007, June). *The validity of logic-based tests*. Paper presented at the 31st Annual Conference of IPMAAC, St. Louis, MO.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 23(5), 565-578.
- Viswesvaran, C., Ones, D. S., Schmidt, F. L., Le, H., & Oh, I. S. (2014). Measurement error obfuscates scientific knowledge: Path to cumulative knowledge requires corrections for unreliability and psychometric meta-analyses. *Industrial and Organizational Psychology*, 7(4), 507-518.

**RECEIVED** 8/10/16 **ACCEPTED** 5/10/17

## ONLINE SUPPLEMENTAL MATERIALS

**Expectancy Values Using Monte Carlo Simulation**

To illustrate the noisiness inherent in the expectancy charts created using raw data, we conducted a Monte Carlo simulation. We simulated data for an observed validity coefficient of .26 (the average uncorrected meta-analytic coefficient for cognitive ability tests reported by Hunter, Schmidt, & Le, 2006) for 10,000 samples consisting of 151 cases (the minimum required for 90% power) each. We then divided the predictor and criterion into quintiles and computed expectancy values using the simulated raw data. Table S1 provides a summary of these results as well as the values that would be obtained using the new methodology that we describe in the paper. We also depict these results in Figure S1a-e, which contains histograms of the five expectancy values for the simulated raw data across the 10,000 samples. The results for the simulated data are shown in blue bars. For reference, we also applied our new approach to computing expectancies to the observed validity coefficients in the 10,000 samples. Histograms for the expectancies from our new approach are shown in Figure S1s-e using red bars. Notice that the distribution of the blue bars (i.e., the expectancies computed using the simulated raw data) are much wider than those for the red bars (i.e., the expectancies computed using our new approach). Overall, the simulated raw data, which necessitate dividing the sample of 151 cases into 25 categories (i.e., 5 predictor score ranges  $\times$  5 criterion score ranges) and computing expectancies on the slices, results in unstable estimates.

As shown in Table S1, for the highest quintile (i.e., the top 20% of scores) on the test, the mean observed expectancy value from the simulated raw was 31.2%; however, the observed values ranged from 0% to 65.2%. Given the SD

of 8.5%, the 95% confidence interval ranged from 14.2% to 48.3%. In contrast, placing a 95% confidence interval around the validity of .26 (i.e.,  $r = .11$  and  $.40$ ) and using our syntax approach yielded values of 24.5% and 38.2%, which is about less than half the size of the observed interval. Furthermore, after correcting for range restriction and criterion unreliability, the meta-analytic validity estimate is .54, which yields an expectancy of 46.0% (substantially higher than the observed mean value of 31.2%).

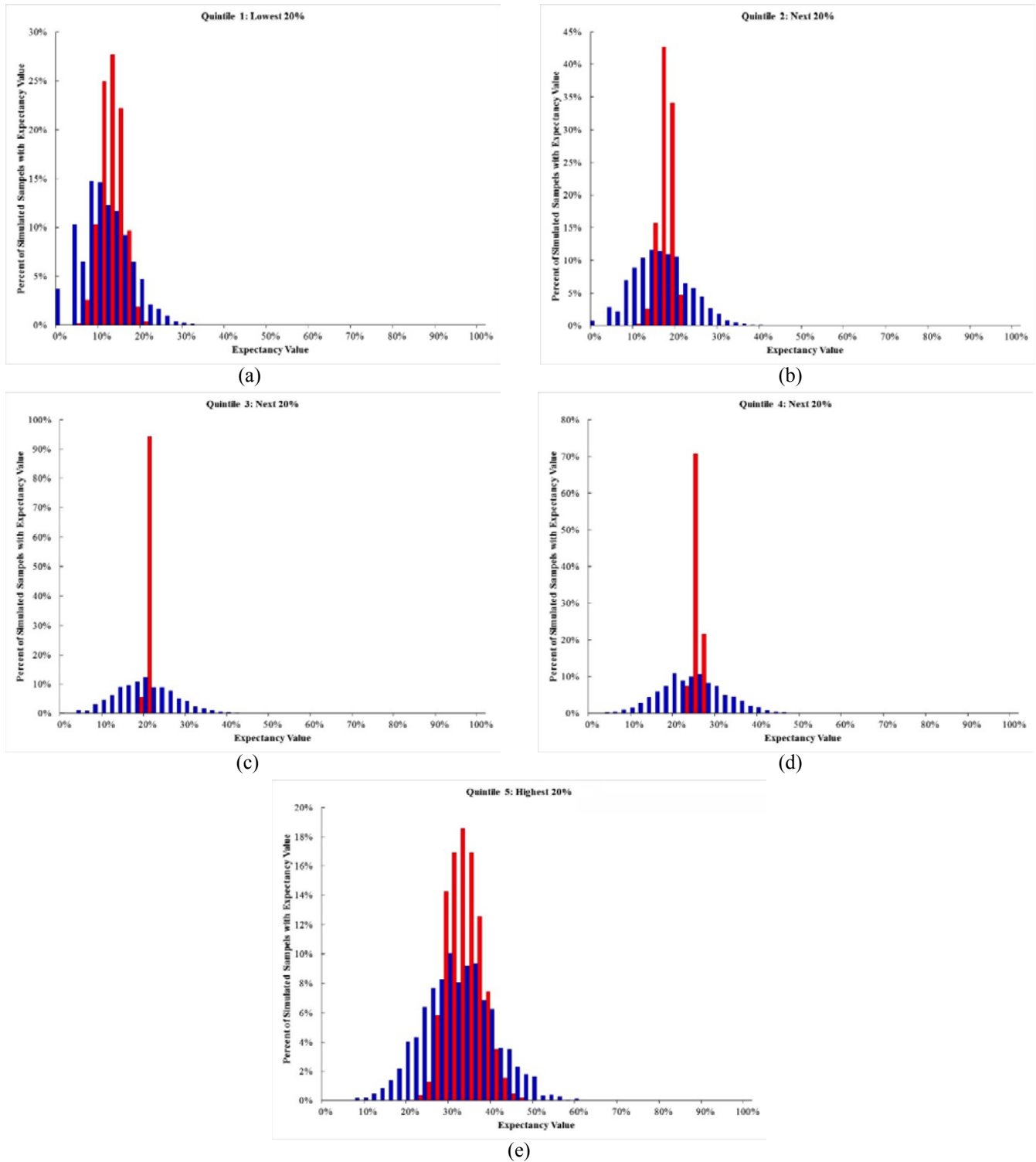
Additionally, we compared the expectancy values for adjacent score ranges on the predictor (e.g., the expectancy values for the lowest quintile and the second lowest quintile were compared). If the predictor and criterion have a linear relationship, the expectancies should have a positive trend going from the lowest test score range to the highest test score range. A total of 87.4% of the samples had at least one pair of adjacent expectancy values (e.g., 13.9% for the lowest quintile and 7.7% for the second lowest quintile) that suggested the relationship between the test and the criterion was non-monotonic. Given past research suggesting that most oft-researched I-O predictors (e.g., cognitive ability tests) are linearly related to performance (with the possible exception of personality measures), the non-monotonicity is an unexpected result.

Finally, we isolated the 485 simulated samples that had an observed validity of exactly .26 (after rounding). The expectancies from these samples also had considerable variability, as shown in Table S1. Thus, even if a researcher is lucky and obtains a point estimate of the validity (.26) that is identical to the true population estimate (.26), there is still considerable variability in the expectancy values (we attribute this variability to sampling error).

**TABLE S1.**

*Comparison of the Results Obtained from Monte Carlo Simulation and the New Method*

|  | Quintile 1<br>Lowest 20% | Quintile 2<br>Next 20% | Quintile 3<br>Next 20% | Quintile 4<br>Next 20% | Quintile 5<br>Highest 20% |
|--|--------------------------|------------------------|------------------------|------------------------|---------------------------|
| True Expectancy $r = .26$  | 10.8%                    | 15.6%                  | 19.2%                  | 23.4%                  | 31.2%                     |
| True Expectancy $r = .11$ (95% confidence interval-lower)            | 15.9%                    | 18.3%                  | 19.9%                  | 21.6%                  | 24.5%                     |
| True Expectancy $r = .40$ (95% confidence interval-upper)            | 6.7%                     | 12.6%                  | 18.0%                  | 24.7%                  | 38.1%                     |
| True Expectancy $\rho = .54$   | 3.4%                     | 9.2%                   | 16.0%                  | 25.6%                  | 46.0%                     |
| Observed Expectancy Across All 10,000 Samples                        |                          |                        |                        |                        |                           |
| <i>M</i>   | 10.7%                    | 15.6%                  | 19.1%                  | 23.4%                  | 31.2%                     |
| <i>SD</i>  | 5.7%                     | 6.7%                   | 7.2%                   | 7.8%                   | 8.5%                      |
| Min  | 0%                       | 0%                     | 0%                     | 0%                     | 0%                        |
| Max  | 40.7%                    | 47.6%                  | 53.3%                  | 56.0%                  | 65.2%                     |
| 95% CI-lower   | -0.8%                    | 2.2%                   | 4.7%                   | 7.9%                   | 14.2%                     |
| 95% CI-upper   | 22.1%                    | 28.9%                  | 33.5%                  | 39.0%                  | 48.3%                     |
| Observed Expectancy For 485 samples with $r_{\text{observed}} = .26$ |                          |                        |                        |                        |                           |
| <i>M</i>   | 10.6%                    | 15.2%                  | 18.6%                  | 23.1%                  | 31.6%                     |
| <i>SD</i>  | 5.5%                     | 6.3%                   | 6.8%                   | 7.9%                   | 7.4%                      |
| Min  | 0%                       | 0%                     | 3.0%                   | 4.2%                   | 10.7%                     |
| Max  | 31.0%                    | 40.7%                  | 38.9%                  | 45.0%                  | 55.2%                     |
| 95% confidence interval-lower  | -0.4%                    | 2.6%                   | 5.0%                   | 7.3%                   | 16.7%                     |
| 95% confidence interval-upper  | 21.6%                    | 27.8%                  | 32.2%                  | 38.9%                  | 46.5%                     |



**FIGURE S1.**

This figure presents histograms of the expectancy values obtained for observed validity coefficients using the raw data approach (shown in blue) and the new approach (shown in red) described in the main paper. Separate sets of histograms are presented for each quintile of test scores.



**Mathematical Theory Behind the Bivariate Normal Distribution**

The new approach to developing expectancy charts described in the paper takes advantage of the bivariate normal distribution. This section provides more information on the mathematical framework behind the bivariate normal distribution.

A bivariate-normal distribution consists of two normally distributed variables,  $x_1$  and  $x_2$ , with a correlation of  $-1 < r < +1$  (see Figure S2a-c). When  $x_1$  and  $x_2$  are uncorrelated (i.e.,  $r_{x_1,x_2} = 0$ ), the distribution consists of a 3-dimensional bell-shaped volume (see Figure S2a). When  $x_1$  and  $x_2$  are correlated, the distribution becomes flatter as the correlation increases (Figure S2b-c). Note that when  $|r_{x_1,x_2}| = 1$ , the bivariate-normal distribution is equivalent to univariate-normal distribution since the two variables,  $x_1$  and  $x_2$ , are perfectly correlated.

It is possible to represent the bivariate normal distribution formulaically. Gatignon (2010) provides the following formula for a bivariate-normal distribution:

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}}$$

Note that the means and standard deviations of  $x_i$  are  $\mu_i$  and  $\sigma_i$  and the correlation between  $x_1$  and  $x_2$  is  $\rho$ . When both variables are standardized on a z-scale, this formula reduces to:

$$\frac{1}{2\pi\sqrt{1-\rho^2}} e^{\left\{-\frac{[x_1^2+x_2^2-2\rho x_1x_2]}{2(1-\rho^2)}\right\}}$$

Integration, from calculus, can be used to find the volume under a function (including 3-dimensional functions). The volume under the normal bivariate distribution is represented by the integral shown below. Note that here, the integration occurs on both variables  $x_1$  and  $x_2$ :

$$\iint \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\left\{-\frac{[x_1^2+x_2^2-2\rho x_1x_2]}{2(1-\rho^2)}\right\}} dx_1 dx_2$$

Vasicek (1998; see also Gupta, 1962) describes the standard process for solving the integral, which is based on the tetrachoric procedure and the Hermite (1864) polynomials.

$$\begin{aligned} & \iint \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\left\{-\frac{[x_1^2+x_2^2-2\rho x_1x_2]}{2(1-\rho^2)}\right\}} dx_1 dx_2 \\ &= \left(\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} dx_1\right) \left(\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} dx_2\right) + \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}\right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}}\right) \\ & \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \left(\sum_{i=0}^{k/2} \frac{k!}{i!(k-2i)!} (-1)^i 2^{-i} x_1^{k-2i}\right) \left(\sum_{i=0}^{k/2} \frac{k!}{i!(k-2i)!} (-1)^i 2^{-i} x_2^{k-2i}\right) \end{aligned}$$

Substituting the Kerridge and Cook’s (1976) approximation for the two integral terms on the left side of the equation produces the following equation:

$$\begin{aligned} &= \left(\frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \left[x_1 + \frac{x_1^3}{3} + \frac{x_1^5}{(3)(5)} + \frac{x_1^{2n+1}}{(2n+1)!} + \dots\right]\right) \times \\ & \left(\frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} \left[x_2 + \frac{x_2^3}{3} + \frac{x_2^5}{(3)(5)} + \frac{x_2^{2n+1}}{(2n+1)!} + \dots\right]\right) + \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}\right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}}\right) \times \\ & \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \left[\left(\sum_{i=0}^{k/2} \frac{k!}{i!(k-2i)!} (-1)^i 2^{-i} x_1^{k-2i}\right) \left(\sum_{i=0}^{k/2} \frac{k!}{i!(k-2i)!} (-1)^i 2^{-i} x_2^{k-2i}\right)\right] \end{aligned}$$

In practice, the volume of the multivariate distributions is computed numerically. Genz and colleagues (Genz et al., 2008; Hothorn, Bretz, & Genz, 2001) developed a package, mvtnorm, for the statistical program R (Ihaka & Gentleman, 1996) that can be used to estimate the area under bivariate and multivariate-normal distributions. One of the algorithms available in the package is MIWA, which uses a numerical methodology to compute the area under the distributions (Miwa, Hayter, & Kuriki, 2003; Mi, Miwa, & Hothorn, 2009). The MIWA procedure can be used when the number of variables in the multivariate-normal distribution is 20 or less<sup>1</sup> (Genz et al., 2008; Mi et al., 2009). Since expectancy charts only require the use of two variables, the variable limitation is not an issue.

Unlike other estimation algorithms in the GenzBretz procedure, Miwa et al.’s (2003) procedure does not rely on Monte Carlo analysis; therefore, it provides consistent results each time it is conducted. Instead of using the equation presented above from Vasicek (1998), Miwa et al. developed a numerical integration approach that can be applied to more than two variables. In their approach, they divide the volume of the multivariate-normal distribution into different sections and compute the volume within each section. Their approach allows a user to modify the size of each section by inputting the number of “grid points” (the default is 128). This type of numerical integration is commonly used in calculus for finding the area under curved functions that are not easily integrated using integration rules from calculus (Larson & Edwards, 2014). The use of numerical integration is not completely foreign to psychometrics; the item-response theory software package BILOG uses numerical integration (Mislevy & Stocking, 1989). Thus, rather than implementing the formulaic integral shown above, Miwa et al.’s procedure implements numerical integration to estimate the volume of a multivariate-normal distribution.

Using 128 grid points, Miwa et al. (2003) tested the accuracy of their approach by comparing results from it to those tabulated by Tong (1990) and Gupta (1963). Of the 567 entries in Tong’s table, Miwa et al.’s procedure agreed

<sup>1</sup> The GenzBretz algorithm can be used for up to 1,000 variables (Genz, Bretz, Miwa, Mi, Leisch, Scheipl, & Hothorn, 2008; Mi, Miwa, & Hothorn, 2009).

with 556 values; the remaining 11 values only disagreed by less than .00002. Miwa et al. noted that their results were similar to those from a Monte Carlo approach; however, their procedure was able to obtain more accurate results with less computing time when the number of variables was less than seven.

Our implementation of Miwa et al.'s (2003) procedure consists of three main computations. First, the joint probability is computed. This is the volume of the bivariate-normal distribution between the cutoffs for the predictor and the criterion. In mathematical terms, where  $x$  is the predictor and  $y$  is the criterion, the joint probability is the intersection ( $\cap$ ) of Lower  $x$  Cutoff  $< x <$  Upper  $x$  Cutoff and Lower  $y$  Cutoff  $< y <$  Upper  $y$  Cutoff. Second, the probability that any given score falls within the cutoffs for

$x$  is computed. This is the univariate volume of the following distribution: Lower  $x$  Cutoff  $< x <$  Upper  $x$  Cutoff. The joint probability is then divided by the  $x$  probability to give the expectancy. Thus, the expectancy is the proportion of those cases within the  $x$  cutoffs that are also within the  $y$  cutoffs. To provide an example, suppose that the validity is zero and we are interested in predictor scores and criterion scores above the mean. Here the  $x$  probability is  $0 < x < \infty$ , which is 0.5. The joint probability ( $0 < x < \infty \cap 0 < y < \infty$ ) is computed to be 0.25 and the expectancy is  $.25 \div .50$ , or 50%. In other words, in total 25% (i.e., a proportion of .25) of the cases have predictor and criterion scores above the mean. Of those with a predictor score above the mean, 50% have a criterion score above the mean. Thus, the expectancy is 50%.

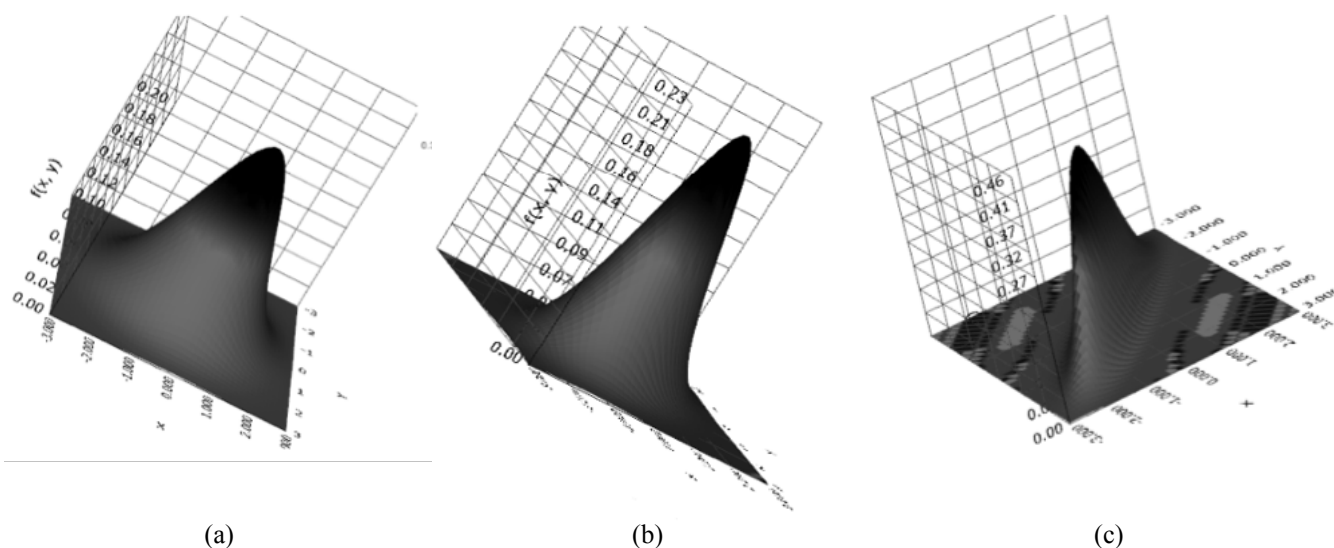


FIGURE S2.

- 3-dimensional plot of a bivariate-normal distribution when  $r_{x1,x2} = 0$ .
- 3-dimensional plot of a bivariate-normal distribution when  $r_{x1,x2} = .5$ .
- 3-dimensional plot of a bivariate-normal distribution when  $r_{x1,x2} = .9$ .

Note. Figures obtained using the Statistics Online Computational Resource (SOCR) web-based statistical program.

## REFERENCES

- Gatignon, H. (2010). *Statistical Analysis of Management Data*. New York, NY: Springer.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2008). mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-2. Available at <http://CRAN.R-project.org/package=mvtnorm>.
- Gupta, S.S. (1962). *Probability integrals of multivariate normal and multivariate t*. (Office of Naval Research Technical Report No. 64). Stanford, CA: Applied Mathematics and Statistics Laboratories, Stanford University.
- Gupta, S.S. (1963). Probability integrals of multivariate normal and multivariate t. *The Annals of Mathematical Statistics*, 34, 792-828.
- Hermite, C. (1864). *Sur un nouveau développement en série des fonctions*. Mallet-Bachelier. [On a new series of expansion functions]. *Comptes Rendus de l'Académie des Sciences Paris* 58 1864 93-100; *Oeuvres* II 293-303.
- Hothorn, T., Bretz, F., & Genz, A. (2001). On multivariate t and Gauss probabilities in R. *sigma*, 1000, 3.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594-612.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Kerridge, D.F., & Cook, G.W. (1976). Yet another series for the normal integral. *Biometrika*, 63, 401-403.
- Larson, R., & Edwards, B. (2014). *Calculus: Early transcendental functions*. Cengage Learning.
- Mi, X., Miwa, T., & Hothorn, T. (2009). mvtnorm:

- New numerical algorithm for multivariate normal probabilities. *The R Journal*, 1(1), 37-39.
- Mislevy, R.J., & Stocking, M.L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement*, 13, 57-75.
- Miwa, T., Hayter, A.J., & Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1), 223-234.
- Tong, Y.L. (1990). *The multivariate normal distribution*. New York, NY: Springer.
- Vasicek, O.A. (1998). A series expansion for the bivariate normal integral, *Journal of Computational Finance*, 1(4), 5-10.