Practical Assessment, Research, and Evaluation

A comparison of two approaches to correction of restriction of range in correlation analysis

Marie Wiberg

Anna Sundström

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

A comparison of two approaches to correction of restriction of range in correlation analysis

Marie Wiberg and Anna Sundström
Umeå University, Sweden

A common problem in predictive validity studies in the educational and psychological fields, e.g. in educational and employment selection, is restriction in range of the predictor variables. There are several methods for correcting correlations for restriction of range. The aim of this paper was to examine the usefulness of two approaches to correcting for range restriction; Thorndike's case 2 correction and ML estimates obtained from the EM algorithm, by comparing the corrected correlations with the correlation from an unrestricted sample. The unrestricted sample consisted of examinees who took the practical Swedish driving-license test regardless of their result on the theory test. Examinees that passed the theory test and took the practical test were regarded as a restricted sample. The result provided empirical support for the appropriateness of Thorndike's case 2 correction method. Although using the EM algorithm yielded a good estimate of the correlation in the unrestricted sample, further studies are needed on this topic.

A common problem in predictive validity studies in educational and psychological research is that the criterion variable is restricted in range. The most common form of range restriction occurs when a researcher wants to estimate the correlation between two variables (x and y) in a population, but subjects are selected on x, and data for y are only available for a selected sample (Raju & Brand, 2003). This occurs for example when scores from admission tests are used to predict academic success in higher education or are compared with grades in the program they were admitted to (Gulliksen, 1950; Thorndike, 1949). Another example is personnel selection, when test scores used for hiring applicants are related to the job performance of those who were hired (Schmidt & Hunter, 1998). Because selection is made on the basis of scores from these kinds of instruments, the range of scores is restricted in the sample. Although the correlation between test scores and academic success or job performance can be obtained for the restricted sample, the correlation for the population of applicants remains unknown. Due to the range

restriction in test scores, the correlation obtained is expected to be an underestimate of the correlation in the population (Hunter & Schmidt, 1990; Henriksson & Wolming, 1998).

The difference in magnitude between correlations in samples with range restriction and the population has been examined in a few studies. For example, in a study described by Thorndike (1949), the relationship between results from an application test used for selection to the American Air Force and subsequent performance results in a test administered during the course was investigated. The results showed only a weak correlation ($\hat{r} = .18$) between test score and education performance, if only those admitted to the program were used in the analysis. In order to investigate the relationship between test results and subsequent performance for the population, all test-takers that took the selection test were allowed to participate in the education, even though their test result was poor. The findings

indicated that when the unrestricted sample was studied, the correlation increased ($\hat{\rho}_{XY}$ = .64).

In most cases, however, researchers do not have access to unrestricted samples. In order to obtain an estimate of the correlation in the population, formulas for corrections of range restriction are commonly applied in predictive validity studies. For example, in validity studies of several large-scale testing programs, such as the Graduate Management Admission Test (GMAT) (Sireci & Talento-Miller, 2006; Talento-Miller & Rudner, 2008), Graduate Record Examinations (GRE) (Chernyshenko & Ones, 1999; Kuncel, Hezlett, & Ones, 2001) and Scholastic Aptitude Test (SAT) (Weitzman, 2005), the validity coefficients have been corrected for range restriction. Correction methods for restriction of range are frequently applied in other settings as well; e.g. in personnel selection when examining the predictive validity of scores from measures of General Mental Ability (Schmidt & Hunter, 1998), in examining the predictive validity of Grade Point Average (GPA) on career development (Cohen-Schotanus et al., 2006) and of tests for selection of officers within the air force (Caretta & Ree, 1995).

Several correction methods for range restrictions have been suggested depending on what kind of corrections are needed and on whether we have a univariate or multivariate data material (Duan & Dunlap, 1997; Held & Foley, 1994; Sackett & Yang, 2000; Theron, 1999; Thorndike, 1949). Which correction methods could be applied also depends on whether range restriction is direct or indirect. In a situation where applicants are selected directly on test scores, the range restriction is *direct*. In a case where the applicants are selected on another variable that is correlated with the test scores, the range restriction is *indirect*. The commonly used Thorndike's case II correction formula is suitable for direct restriction of range. This formula has however been applied for indirect restriction of range even though it has been shown to underestimate validity coefficients. Lately, a formula for indirect range restriction has been developed (Schmidt, Oh, & Le, 2006). In the present study we have a case of direct range restriction. Therefore this latter formula will not be applied in the present study. Although several methods for range restriction exist, these methods are not flawless, since the use of these methods can result in either undercorrection or overcorrection of the observed

relationship (Linn, 1967). Further, the accuracy of for example estimating standard errors can differ among various methods (Duan & Dunlap, 1997).

In this article, the correlations in a restricted and an unrestricted sample will be examined using two different approaches for correction of direct range restriction. This study is unique because it is rare to have the opportunity to study empirically the correlations in a range restricted sample compared with an unrestricted sample.

The first approach used in this paper is a common correction formula that is available for correcting correlation estimates for restriction of range (Hunter & Schmidt, 1990). The formula was originally developed by Karl Pearson (1903) but became widely known through the work of Thorndike (1949). The formula is known as *Thorndike case 2* and has been shown to produce close estimates of the correlation in a population. For example, this formula was used by Chernyshenko and Ones (1999) when examining the correlation between GRE scores and GPA in a restricted sample. The correlations in the restricted sample ranged between $\hat{r}$ = .15 and $\hat{r}$ = .37 and when the correction formula was applied, the correlations were stronger and ranged between $\hat{r}$ = .35 and $\hat{r}$ = .70.

The second approach used in this paper is seldom used with range restriction problems, although it has been mentioned as a possibility (Mendoza, 1993). Using this approach we will view the selection mechanism as a missing data mechanism, i.e. we will view the data in the variable that is restricted in range as missing, and estimate the missing values before estimating the correlation. By viewing it as a special case of missing data, we can borrow from a rich body of statistical methods; for an overview see e.g. Little & Rubin (2002), Little (1992) or Schafer & Graham (2002). There are three general missing data situations; MCAR, MAR and MNAR. Assume X is a variable that is known for all examinees and Y is the variable of interest with missing values for some examinees. MCAR means that the data is Missing Completely At Random, i.e. the missing data distribution does not depend on the observed or missing values. In other words, the probability of missingness in data Y is unrelated to X and Y. MAR means that the data is Missing At Random, i.e. the conditional distribution of data being missing given

2

the observed and missing values depends only on the observed values and *not* on the missing values. In other words, the probability of missingness in data Y is related to X, but not to Y. MNAR means that data is Missing Not At Random. In other words, the probability of missingness on Y is related to the unobserved values of Y (Little & Rubin, 2002; Schafer & Graham, 2002). If the data is either MCAR or MAR, we can use imputation methods to replace missing data with estimates. In this paper we have a selection mechanism that is based solely on X, hence we will consider the data to be MAR, which is in line with Mendoza's (1993) conclusions about similar data. In this approach, we can use information on some of the other variables to impute new values. Herzog & Rubin (1983) stated that by using imputation one can apply existing analysis tools to any dataset with missing observations and use the same structure and output.

There are several different techniques that use imputation to replace missing values. The most commonly applied techniques are mean imputation, hot-deck imputation, cold-deck imputation, regression imputation and multiple imputations (Madow, Olkin, & Rubin, 1983; Särndal, Swensson, & Wretman, 1992). In general, imputation may cause distortions in the distribution of a study variable or in the relationship between two or more variables. This disadvantage can be diminished when e.g. multiple regression imputation is used (Särndal et al., 1992). For example, Gustafsson & Reuterberg (2000) used regression to impute missing values in order to get a more realistic view of the relationship between grades in upper secondary schools in Sweden and the Swedish Scholastic Achievement Test. Note that regression imputation is questionable to use, because all imputed values fall directly on the regression line, the imputed data lack variability that would be present had both X and Y been collected. In other words the correlation would be 1.0 if only computed with imputed values (Little & Rubin, 2002). Therefore we suggest using imputed Maximum Likelihood (ML) estimates for the missing values that are obtained using the Expectation Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977).

Aim

The aim was to compare two methods for range restriction correction; Thorndike's case 2 correction method and ML estimates obtained from the EM

algorithm, to the correlation in an unrestricted sample.

The data used in the present study was results from the Swedish driving-license theory and practical test. In both Sweden and Great Britain, attempts have been made to investigate the relationship between examinees' results on the theory test and the practical test. One problem that these studies have in common is restriction in range in theory test scores, because examinees have to pass the theory test before being allowed to take the practical test (Forsyth, 1992; Wolming & Wiberg, 2004). Studies that have explored the relationship between the Swedish theory and practical tests (Sundström, 2003; Wiberg, 2004; Wolming, 2000; Wolming & Wiberg, 2004) indicate that there is a weak negative relationship between the scores in the theory test and the number of competence shortages in the practical test, i.e. good results on the theory test are related to few competence shortages in the practical test. The weak correlation between the tests has been explained as a result of the range restriction in the theory test scores. It is important to examine the relationship between the theory test and practical test, because the two tests are viewed as parts of a driving-license examination. This means that the theoretical knowledge examined in the theory test is important in the practical driving, but the practical driving skills are not critical for mastering the theoretical content. Therefore we hypothesise that there should be a moderate relationship between these tests.

DATA AND METHOD

Instruments

The Swedish driving-licence test consists of a theory test and a practical test. The theory test is a computerized criterion-referenced mastery test that consists of 65 dichotomously scored multiple-choice items. The examinees have to achieve a total score of at least 52 (i.e. 80 percent) to pass the test. The internal consistency was assessed through Cronbach's alpha. For the sample used in the present study, $\alpha = 0.76$. In the practical test, the examinee's driving performance in different traffic situations is assessed by a driver examiner with respect to five competences that are related to the driver's awareness of risks in traffic. The examiner uses a special form to record what has been tested and if the examinee has failed with respect to any of the competences. If the

examinee fails in any competence he or she fails the test (Vägverket, 1996).

Participants

The sample used was a large random sample that was representative of the population of driving-license examinees in Sweden. This sample was divided into an "unrestricted sample" and a "restricted sample". The unrestricted sample included 2,254 examinees, 1,313 (58 %) men and 941 (42 %) women that either passed or failed the theory test and then took the practical test. Their average age was 21.5 years and ranged between 18 and 69 years (SD = 7.1, Mdn = 19). The pass rate was 68.1 percent on the theory test and 61.3 percent on the practical test. The restricted sample included those examinees that passed the theory test, i.e. 1,535 examinees. Of these 852 (55 %) were men and 683 (45 %) were women. Their average age was 20.7 years and ranged between 18 and 64 years (SD = 2.6, Mdn = 18). In the restricted sample 69.6 percent of the examinees passed the practical test.

Procedure

The unrestricted and restricted sample included examinees who participated in a project conducted by the Swedish Road Administration during six months in 2006, when a new model for the driving test was tested. The study was conducted at six driving test centres in three regions in Sweden. The overall test results in these regions have been shown to be representative of the population of examinees who take the Swedish driving-license test (Wiberg, Stenlund, Sundström, & Henriksson, 2005). In the project, examinees were given the opportunity to take the practical test regardless of their result on the theory test. Usually only examinees who pass the theory test are allowed to take the practical test, i.e. there is a restriction of range. In order to resemble the usual procedure, those who passed the theory test and took the practical test were viewed as the restricted sample.

Analysis

In order to examine the relationship between the results of the theory and practical tests, Pearson's product-moment correlation, denoted *r*, was used together with scatter-plots. The relationship between the results from the theory and practical tests is assumed to be linear, which our analysis supported since no curvilinear relationship could be detected.

Note that we examined the material for possible effects of outliers. Since we arrived at the same results with and without outliers included, and since this is a real data set, we decided to keep all data in the analyses.

In order to deal with range restriction (in the restricted sample), two different approaches were used and the results from them were compared with the estimated correlation obtained from the unrestricted sample, denoted $\hat{\rho}_{XY}$. SPSS 14.0 software was used to calculate observed and corrected correlations in both samples. Firstly, Karl Pearson's (1903) formula, which is usually referred to as Thorndike (1949) case 2 for explicit selection, was used. Explicit selection means that there is a direct selection on X, i.e. no one with a test score below a specified cutscore on X is selected. In the present study, X refers to the score on the theory test. Gross & McGanney (1987) claim that this selection process is ignorable. It requires that the unrestricted variance is known for the selection variable and that there is no additional range restriction on additional variables (Sackett & Yang, 2000). This formula has been widely discussed and used (Chernyshenko & Ones, 1999; Gross & McGanney, 1987; Holmes, 1990; Mendoza, 1993; Sackett & Yang, 2000). The formula uses the correlation of the restricted sample and the standard deviation of the independent variable (X) in the restricted sample and in the unrestricted sample to provide an estimate of the correlation in the population:

$$r_{XY} = S_X r_{xy} / (S_X^2 r_{xy}^2 + s_x^2 - s_x^2 r_{xy}^2)^{1/2} \qquad (1)$$

where

r_{xy} is the observed correlation between X and Y in the restricted sample.

s_x is the estimated standard deviation of X in the restricted sample.

S_X is the estimated standard deviation of X in the unrestricted sample.

r_{XY} is the estimated corrected correlation between X and Y in the unrestricted sample when only the restricted sample has been used.

Formula (1) has been shown to give a close estimate of the true correlation (ρ_{XY}) if the regression

of Y on X is linear and homoscedastic, i.e. the variance of the error term is the same in the restricted sample and in the population (Gulliksen, 1950; Levin, 1972; Lord & Novick, 1968). Bivariate normality is a sufficient although not a required condition for this formula (Lawley, 1943). Gross (1982) and Gross & Fleischman (1983) showed that these assumptions can be relaxed in many circumstances. E.g. even if the regression is nonlinear and heteroscedastic, the corrected correlation will be a more accurate estimate than the uncorrected correlation (Chernyshenko & Ones, 1999).

Secondly, EM imputation was used to obtain estimated values on pseudo missing values on the practical test. In other words, we constructed a range-restricted sample by removing the practical test results among those examinees who failed the theory test, therefore these observations were viewed as missing values. Using the definition of missing data in the introduction the data are assumed to be MAR since the probability of missingness in the practical test score (Y) is related to the theory test score (X), but not to the practical test score (Y). Maximum likelihood (ML) estimates using the Expectation Maximization (EM) algorithm were imputed for the practical test for examinees who failed the theory test (Dempster et al., 1977; Little, 1992). The complete and incomplete cases were used together as the EM algorithm reestimates means, variances and covariances until the process converges. The base of EM missing values is an iterative regression imputation. The final estimated moments are the EM estimates including estimates for the correlation. For an extensive description see SPSS (2002). Dempster, Laird & Rubin (1977) showed that these are maximum likelihood estimates, which are consistent, i.e. they converge in probability to the population parameters. The idea is that the missing Y values are imputed using

$$Y_{imp} = \hat{\alpha}_{EM} + \hat{\beta}_{EM} X \ ,$$

where $\hat{\alpha}_{EM}$ and $\hat{\beta}_{EM}$ are the estimates obtained from the final iteration of the EM algorithm.

When maximum likelihood estimates are obtained using the EM algorithm it is denoted $\hat{r}_{EM}$. Schaffer (2002) suggested that using EM imputation is valid when examining missing data.

RESULTS

In the first step, examination of the unrestricted sample indicated that there was a significant negative relationship between results on the theory and practical tests ($\hat{\rho}_{XY} = -.28$). The restricted sample demonstrated a weaker correlation between theory and practical test results ($\hat{r}_{xy} = -.12$) (see Figure 1).
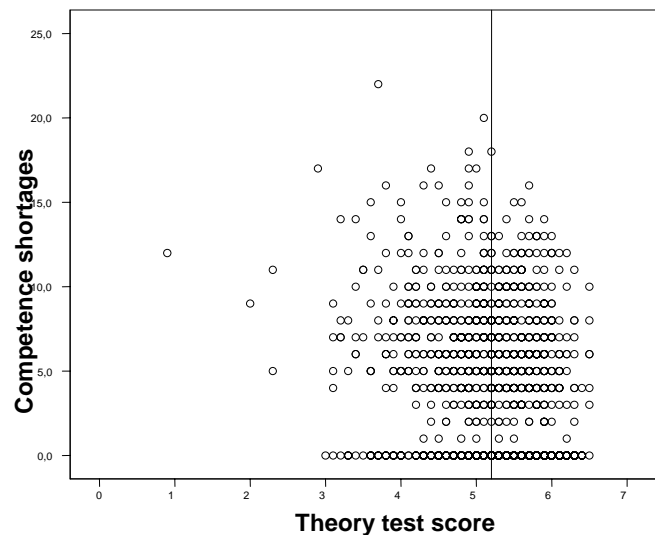
Figure 1. Plot of the competence shortages and the theory test scores in the unrestricted sample. The vertical line indicates the cut-score of 52 on the theory test.

In the next step, the two approaches for dealing with range restriction were applied to the restricted sample. The first approach was to apply the correction formula (1) to the correlation in the restricted sample, which resulted in an estimated corrected correlation of $\hat{r}_{XY} = -.24$. The second approach was to replace the pseudo missing values with ML estimates using the EM algorithm, which resulted in an estimated correlation of $\hat{r}_{EM} = -.28$ (see Table 1).

DISCUSSION

A common methodological problem in test validation studies is restriction of range on the predictor due to explicit selection. There are several methods that can be used to correct correlations for restriction of range, and some have been more frequently used than others. In this study we had a unique opportunity to examine empirically the usefulness of two different approaches by comparing the estimated correlations with the correlation from an unrestricted sample. The correlation obtained in the restricted sample ($r = -.12$) was similar, with respect to strength and direction, to

correlations obtained in previous studies of the Swedish driving-license test (Sundström, 2003; Wiberg, 2004). There was a negative correlation between the results on the theory and practical tests in both the restricted and unrestricted sample. The reason why the correlation was negative is that scores on the theory test were correlated with competence shortages in the practical test. In the theory test the better performance, the higher the score, and on the contrary, in the practical test the worse the

performance, the higher the score. Due to range restriction, the correlation in the restricted sample was weaker. The estimates obtained using the correction methods as well as the estimated population correlation indicated that there is a moderate correlation between the theory and practical tests. This supports the notion that theoretical knowledge and practical driving performance are integrated and that the tests can be viewed as two parts of one driving-license examination.

Table 1. Standard deviations for the restricted (s_x) and unrestricted (S_X) sample, correlation estimates in the restricted sample ($\hat{r}_{xy}$), correlation estimates using the correction formula 1 ($\hat{r}_{XY}$), ML estimate ($\hat{r}_{EM}$), and correlation estimate in the unrestricted sample ($\hat{\rho}_{XY}$).

	s_x	S_X	$\hat{r}_{xy}$	$\hat{r}_{XY}$	$\hat{r}_{EM}$	$\hat{\rho}_{XY}$
Estimated correlations	3.05	6.28	-0.12**	-0.24	-0.28**	-0.28**

* p < 0.05 ** p < 0.01
Note: Unrestricted sample n = 2254, Restricted sample n = 1535.

The formula for correction of range restriction provided a good estimate of the correlation in the unrestricted sample ($\hat{r}_{XY}$ = -.24). This formula has been used for correcting correlations for range restriction, for example by Sireci and Talento-Miller (2006), where the relationship between GMAT scores and GPA were investigated. In that study, the correlations in the restricted sample were weak and positive (r = .14 - .30) and the corrected correlations were still positive and stronger (r = .19 - .60). However, no information about the correlation in the unrestricted sample was available. The results from the present study also indicated that the correlation using ML estimates obtained from the EM algorithm provided a very good estimate of the correlation in the unrestricted sample. However, because this approach is not commonly used in range restriction studies, the appropriateness and effectiveness of this method should be further examined, e.g. through simulation studies.

Although it is a common problem, range restriction has not been empirically studied in many educational or psychological settings. The main reason for the lack of empirical studies of this topic is that access is rare to unrestricted samples to compare the corrected correlations with. As mentioned above, this study compared the estimated correlations

obtained from two methods for range restriction correction to the correlation in an unrestricted sample. Regarding the first correction method, Thorndike's Case 2, the result confirms previous findings which indicate that the correction method provided a good estimate of the population correlation ($\hat{r}_{XY} = -.24$). Regarding the second correction method, the results indicate that ML estimates obtained from the EM algorithm seem to be a very effective method of estimating the population correlation ($\hat{r}_{EM} = -.28$). Because this approach has not been commonly used in restriction of range studies, its appropriateness for this use needs to be further explored.

Using an appropriate method for correcting for restriction of range is most important when conducting predictive validity studies of instruments used for example for selection to higher education and employment selection. The use of inappropriate methods for range restriction correction or no correction method at all could result in invalid conclusions about test quality. Thus, carefully considering methods for correcting for restriction of range in correlation studies is an important validity issue.

There are still many questions that need to be answered with regard to range restriction. Some areas

of future research that are of special interest include simulations of different population correlations and different selection proportions when using the described missing data approach. Another topic could of course be to examine the approaches if we have a curvilinear relationship instead of a linear relationship between the variables. Regarding the EM imputation approach, one important research question is also how many cases can be imputed at the same time as we obtain a good estimate of the population correlation. This would also be an excellent topic for future studies.

REFERENCES

Caretta, T. R., & Ree, M. J. (1995). Air force officer qualifying test validity for predicting pilot training performance. *Journal of Business and Psychology, 9*(4), 379-388.

Chernyshenko, O. S., & Ones, D. S. (1999). How selective are psychology graduate programs? The effect of the selection ratio on GRE score validity. *Educational and Psychological Measurement, 59*(6), 951-961.

Cohen-Schotanus, J., Muijtjens, A. M. M., Reinders, J. J., Agsteribbe, J., Rossum, H. J. M., & van der Vleuten, C. P. M. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education, 40*, 1012-1019.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*(39), 1-38.

Duan, B., & Dunlap, W. P. (1997). The accuracy of different methods for estimating the standard error of correlations corrected for range restriction. *Educational and Psychological Measurement, 57*, 254-265.

Forsyth, E. (1992). *Cohort study of learner and novice drivers part 2: attitudes, opinions and the development of driving skills in the first 2 years* (Research report No. 372). Crowthorne, Berkshire: Transport research laboratory.

Gross, A. L. (1982). Relaxing assumptions underlying corrections for range restriction. *Educational and Psychological Measurement, 42,* 795-801.

Gross, A. L., & Fleischman, L. (1983). Restriction of range corrections when both distribution and selection assumptions is violated. *Applied Psychological Measurement*, 7, 227-237.

Gross, A. L., & McGanney, M. L. (1987). The restriction of range problem and nonignorable selection processes. *Journal of Applied Psychology, 72*, 604-610.

Gulliksen, H. (1950). *Theory of mental tests.* New York: John Wiley & sons.

Gustafsson, J.-E., & Reuterberg, S.-E. (2000). Metodproblem vid studier av högskoleprovets prognosförmåga - och deras lösning! [Methodological problems associated with the study of the predictive validity of the SweSAT] *Pedagogisk Forskning i Sverige, 5*(4), 273-283.

Held, J. D., & Foley, P. P. (1994). Explanations for accuracy of the general multivariate formulas in correction for range restriction. *Applied Psychological Measurement, 18*, 355-367.

Henriksson, W., & Wolming, S. (1998). Academic performance in four study programmes: a comparison of students admitted on the basis of GPA and SweSAT scores, with and without credits for work experience. *Scandinavian Journal of Educational Research, 42*, 135-150.

Herzog, T. N., & Rubin, D. B. (1983). Using multiple imputations to handle nonresponse in sample surveys. In W. G. Madow, I. Olkin & D. B. Rubin (Eds.), *Incomplete data in sample surveys.* London: Academic Press Inc.

Holmes, D. J. (1990). The robustness of the usual correction for restriction in range due to explicit selection. *Psychometrika, 55*, 19-32.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings.* London: Sage Publications.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations: Implications for Graduate Student Selection and Performance. *Psychological Bulletin, 127*(1), 162-181.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinborough, LXIII-Part 1*, 28-30.

Levin, J. (1972). An occurence of an increase in correlation by restriction of range. *Psychometrika, 37*, 93-97.

Linn, R. L. (1967). *Range restriction problems in the validation of a guidance test battery.* Princeton, NJ: Educational Testing Service.

Little, J. A. (1992). Regression with missing X's: a review. *Journal of American Statistical Association, 87*, 1227-1237.

Little, J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2 ed.). Hoboken, NJ: John Wiley and Sons.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Madow, W. G., Olkin, I., & Rubin, D. B. (1983). *Incomplete data in sample surveys.* London: Academic Press Inc.

Mendoza, J. L. (1993). Fisher transformations and correlations corrected for selection and missing data. *Psychometrika, 58*, 601-615.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: II. On the influence of natural selection on the variability and correlation of organs. *Royal Society Philosophical Transactions, 200*(Series A), 1-66.

Raju, N. S., & Brand, P. A. (2003). Determining the Significance of Correlations Corrected for Unreliability and Range Restriction. *Applied Psychological Measurement, 27*(1), 52-71.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: an expanded typology. *Journal of Applied Psychology, 85*, 112-118.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147-177.

Scheffer, J. (2002). Dealing with missing data. *Res Lett Inf Math Sci, 3*, 153-160.

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*(2), 262-274.

Sireci, S. G., & Talento-Miller, E. (2006). Evaluating the Predictive Validity of Graduate Management Admission Test Scores. *Educational and Psychological Measurement, 66*, 305-317.

SPSS. (2002). *MVA*. Available from www.spss.com.

Vägverket [Swedish Road Administration]. (1996). *VVFS 1996:168 Vägverkets författningssamling. vägverkets föreskrifter om kursplaner, behörighet B. [Regulations concerning driving license class B]*. Borlänge: Swedish Road Administration.

Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: implications for selection procedure validities and other research results. *Personnel Psychology, 59,* 281-305.

Sundström, A. (2003). *Den svenska förarprövningen. Sambandet mellan kunskapsprovet och körprovet, provens struktur samt körkortsutbildningens betydelse [Driver testing in Sweden. A study of the relationship between the theoretical and practical test, the structure of the tests and the effect of driver education on test performance]* (PM No. 183). Umeå universitet: Enheten för pedagogiska mätningar.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York Springer-Verlag.

Talento-Miller, E., & Rudner, L. M. (2008). The Validity of Graduate Management Admission Test Scores. A Summary of Studies Conducted From 1997 to 2004. *Educational and Psychological Measurement, 68*(1), 129-138.

Theron, C. C. (1999). The effect of a joint correction for the attenuating effect of criterion unreliability and case 1 restriction of range on expected criterion performance and the conditional probability of success. *Educational and Psychological Measurement, 59*, 787-803.

Thorndike, R. L. (1949). *Personnel selection: test and measurement techniques*. New York: Wiley.

Weitzman, R. A. (2005). The prediction of college achievement by the scholastic aptitude test and the high school record. *Journal of Educational Measurement, 19*(3), 179-191.

Wiberg, M. (2004). *Klassisk och modern testteori. Analys av det teoretiska och praktiska körkortsprovet [Classical and modern test theory: analysis of the theoretical and practical driving-license test]* (BVM No. 5). Umeå universitet: Institutionen för beteendevetenskapliga mätningar.

Wiberg, M., Stenlund, T., Sundström, A., & Henriksson, W. (2005). *Underlag för urval till försöksverksamhet för kursplan B [Considerations for sampling in an experimental project about a new model for driving-license examination]*. Umeå universitet: Institutionen för beteendevetenskapliga mätningar.

Wolming, S. (2000). *Förarprövningens struktur och resultat. En studie av relationen mellan kunskapsprov och körprov samt utbildningsbakgrundens betydelse [The structure and results of driving-license examination]* (PM No. 166). Umeå universitet: Enheten för pedagogiska mätningar.

Wolming, S., & Wiberg, M. (2004). The Swedish driver licensure examination: exploration of a two-stage model. *Journal of Safety Research, 35*, 491-495.

Note

Citation

Corresponding Author

Marie Wiberg
Dept of Educational Measurement
Umeå University
SE- 901 87 UMEÅ
SWEDEN

Tel: +46 90 786 69 43
Fax: + 46 90 786 66 86

E-mail: marie.wiberg [at] edmeas.umu.se