Multiple Linear Regression Viewpoints

A Publication sponsored by the American Educational Research Association's Special Interest Group on Multiple Linear Regression: The General Linear Model

MLRV

Volume 26 • Number 1 • Spring 2000

Table of Contents

A Fixed Effects Panel Data Model: Mathematics Achievement in the U.S.

Todd Sherron **Jeff M. Allen**
University of North Texas

Statistical models that combine cross section and time series data offer analysis and interpretation advantages over separate cross section or time series data analyses (Mátyás & Severstre, 1996). Time series and cross section designs have not been commonplace in the research community until the last 25 years (Tieslau, 1999). In this study, a fixed effects panel data model is applied to the National Education Longitudinal Study of 1988 (NELS:88) data to determine if educational process variables, teacher emphasis, student self-concept, and socio-economic status can account for variance in student mathematical achievement. A model that includes seven independent variables accounted for 25% of the variance in student mathematical achievement test score. The study provides educational researchers with an applied model for panel data analysis.

Time series and cross section designs have not been commonplace in the research community until the last 25 years (Tieslau, 1999). In fact, the U.S. Department of Education's National Center for Education Statistics (NCES) was not mandated to "collect and disseminate statistics and other data related to education in the United States" until the Education Amendments of 1974 (Public Law 93-380, Title V, Section 501, amending Part A of the General Education Provisions Act). Researchers commonly have termed data that contains time series and cross section units to be panel or longitudinal data. In this study, these terms are used interchangeable. Essentially, panel data is a set of individuals who are repeatedly sampled at different intervals in time, across a multitude of cross sectional variables. The term "individual" might be used loosely to imply a person, a household, a school, school districts, firms, or a geographical region. Figure 1 provides a typical Panel data structure. Schools have been used to represent the different "individuals". (Note: the individual unit could just as well have been different schools within a particular district, school districts within a state, or an aggregate representation by state).

Researchers who are interested in understanding, explaining, or predicting variation within longitudinal data are faced with complex stochastic specifications. The problem that occurs when measures exhibit two-dimensional variation—variation across time and cross section, in model specification. In other words, researchers need to specify a model that can capture individual differences in behavior across individuals and/or through time for estimation and inference purposes (Greene, 1997). In general, longitudinal (panel) data sets contain a large number of cross-section units and a relatively small number of time-series units.

The U.S. Department of Education began collecting data in 1988 about critical transitions experienced by students as they leave elementary schools and progress through high school and into postsecondary institutions or the work force. The National Education Longitudinal Study of 1988 (NELS:88) contains data about educational processes and outcomes pertaining to student learning, predictors of dropping out, and school effects on students' access to programs and equal opportunities to learn. The first follow-up was conducted with the same students, their teachers, and principals in 1990. The second follow up survey was conducted in 1992, and the third in 1994. Data from NELS:88 will be used in this study to determine if student perception of educational process variables can account for the variance in mathematical achievement.

Model Specification

When should a fixed effects or random effects model be utilized? The answer to this question is often debated. Some believe that it is dependent upon the underlying cause in the model. For example, if the individual effects are the result of a large number of non-observable stochastic variables, then the random effect interpretation is demanded. Others think the decision rests on the nature of the sample – that is when the sample is comprehensive or exhaustive, then fixed effects models are the natural choice to enhance the generalizability. On the contrary, if the sample does not contain a large percent of the population then the random effects model would be the model of choice. According to Hasiao (1985), it is ultimately, " up to the investigator to decide whether he wants to make an inference with respect to population characteristics or only with respect to effects that are in the sample" (*p*. 131). It is unlikely that this debate will ever be resolved per se, however, if the choice between the two underlying methods is clear, then the estimation method should be chosen accordingly. However, if the choice is not clear, then the decision should be based on the nature of the sample and statistical evidence. For example, if the individual effects are significant then this is a sign that a significant component of the model is accounted for

$$
Y_{it} = \begin{bmatrix} Y_{\text{school\#1, year\#1}} \\ Y_{\text{school\#1, year\#2}} \\ \vdots \\ Y_{\text{school\#1, year\#}T} \\ \cdots \\ Y_{\text{school\#2, year\#1}} \\ Y_{\text{school\#2, year\#2}} \\ \vdots \\ Y_{\text{school\#2, year\#}T} \\ \cdots \\ Y_{\text{school\#}N\text{, year\#1}} \\ Y_{\text{school\#}N\text{, year\#2}} \\ \vdots \\ Y_{\text{school\#}N\text{, year\#}T} \end{bmatrix}
\quad
\boldsymbol{X}_{it} = \begin{bmatrix}
X1_{\text{school\#1, year\#1}} & X2_{\text{school\#1, year\#1}} & XK_{\text{school\#1, year\#1}} \\
X1_{\text{school\#1, year\#2}} & X2_{\text{school\#1, year\#2}} & XK_{\text{school\#1, year\#2}} \\
\vdots & \vdots & \vdots \\
X1_{\text{school\#1, year\#}T} & X2_{\text{school\#1, year\#}T} & XK_{\text{school\#1, year\#}T} \\
\cdots & \cdots & \cdots \\
X1_{\text{school\#2, year\#1}} & X2_{\text{school\#2, year\#1}} & XK_{\text{school\#2, year\#1}} \\
X1_{\text{school\#2, year\#2}} & X2_{\text{school\#2, year\#2}} & XK_{\text{school\#1, year\#2}} \\
\vdots & \vdots & \vdots \\
X1_{\text{school\#2, year\#}T} & X2_{\text{school\#2, year\#}T} & XK_{\text{school\#1, year\#}T} \\
\cdots & \cdots & \cdots \\
X1_{\text{school\#}N\text{, year\#1}} & X2_{\text{school\#}N\text{, year\#1}} & XK_{\text{school\#}N\text{, year\#1}} \\
X1_{\text{school\#}N\text{, year\#2}} & X2_{\text{school\#}N\text{, year\#2}} & XK_{\text{school\#}N\text{ year\#2}} \\
\vdots & \vdots & \vdots \\
X1_{\text{school\#}N\text{, year\#}T} & X2_{\text{school\#}N\text{, year\#}T} & XK_{\text{school\#}N\text{, year\#}T}
\end{bmatrix}
$$

Figure 1. Panel Data Structure

by the individual effects parameter and so fixed effects might be preferred over random effects. However, the Hausman test statistic, a statistic designed to test model fit, can be used to determine when a random effects model is preferred, i.e. a large Hausman test statistic indicates a random effects components (Greene, 1997).

The Fixed-Effects Model

The fixed effects (FE) model takes α_i to be a group specific constant term in the regression equation

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + \varepsilon_{it} \quad (1)$$

or in matrix notation

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}'\beta + \varepsilon_{it} \qquad (2)$$

where $\boldsymbol{X}_{it}' = [X_{1it}, X_{2it}, \ldots, X_{Kit}]$
and $\beta' = [\beta_1, \beta_2, \ldots, \beta_K]$.

The "*i*" indexes cross-section realizations so that $i = 1,2,3, \ldots, N$ and "*t*" indexes time-series realizations so that $t = 1, 2,3, \ldots, T$. The *individual effect* α_i, is regarding to be constant over time (*t*) and specific to the individual cross-sectional unit (*i*). The term α_i is presumed to capture the unobservable, and non-measurable characteristics that differentiate individual units. Basically, this implies that all behavioral differences between individuals (e.g., schools in Figure 1) are fixed over time and are represented as parametric shifts of the regression function. Mátayás and Sevestre (1996, *p*. 34) state, "the intercept is allowed to vary from individual to individual while the slope parameters are assumed to be constants in both the individual and time dimensions".

The fundamental assumption of the fixed effects model are:

$E[\varepsilon_{it}] = 0$,

$Cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$,

$Var(\varepsilon_{it}) = E[\varepsilon^2_{it}] = \sigma^2_e$,

$E[\varepsilon_{it}, X_{1it}] = E[\varepsilon_{it}, X_{2it}] = \ldots = E[\varepsilon_{it}, X_{Kit}] = 0$,
and X_{kit} is not invariant.

Under these assumptions, the ordinary least squared estimator (OLS) can be use to obtain unbiased, consistent, and efficient (BLUE) parameter estimates.

The Random Effects Model

The random effects (RE) model—also know as the error component model, includes a non-measurable stochastic variable, which differentiates individuals. It is written as:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + u_i + \varepsilon_{it} \qquad (3)$$

or in matrix notation

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}'\beta + u_i + \varepsilon_{it} \qquad (4)$$

where $\boldsymbol{X}_{it}' = [X_{1it}, X_{2it}, \ldots, X_{Kit}]$

and $\beta' = [\beta_1, \beta_2, \ldots, \beta_K]$.

The "*i*" indexes cross-section realizations so that $i=1,2,3,\ldots, N$ and "*t*" indexes time-series realizations so that $t = 1,2,3,\ldots, T$. The term "u_i" is a stochastic variable that embodies the *unobservable* or *non-measurable* disturbances that accounts for individual differences. Essentially, the effect is thought to be a random individual effect rather that fixed parameter. For example, a researcher might try to discern whether there is difference in achievement between districts in the State of Texas. Instead of including every school district in the equation (as we would have in the fixed effects model using dummy variables) one can randomly sample school districts and assumes that the effect is random distributed across "individuals" but constant through time.

The fundamental assumptions of the random effects model are as such:

$E[u_i, X_{1it}] = E[u_i, X_{2it}] = . . . = E[u_i, X_{Kit}] = 0,$

$E[\varepsilon_{it}] = [u_i] = 0,$

$Var\ (u_i) = E[u^2_i] = \sigma^2_u$, and

$Cov(u_i, \varepsilon_{it}) = E[u_i, \varepsilon_{it}] = \sigma_{\varepsilon,u},$

Assuming normality $u_i \sim N(0, \sigma^2_u)$, $\varepsilon_{it} \sim N(0, \sigma^2_e)$, both "$u_i$" and "$\varepsilon_{it}$" are stochastic variables, but form one composite error term-called omega $(u_i + \varepsilon_{it}) \equiv \omega_{it}$,

$$\text{where} \quad \omega_{it} = \begin{bmatrix} u_1 + \varepsilon_{11} \\ u_1 + \varepsilon_{12} \\ \vdots \\ u_1 + \varepsilon_{1T} \\ \cdots \\ u_N + \varepsilon_{N1} \\ u_N + \varepsilon_{N2} \\ \vdots \\ u_N + \varepsilon_{NT} \end{bmatrix}$$

The error term now consist of two components: (1) the error disturbance ε_{it}, and (2) the individual specific disturbance u_i,. The RE model now takes the form of

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} \ldots + \beta_K X_{Kit} + \omega_{it} \quad (5)$$

or in matrix notation

$$Y_{it} = \alpha + \boldsymbol{X}_{it}'\beta + \omega_{it} \qquad (6)$$

The error term in the model now exhibits the following characteristics:

$$Var(\omega_{it}) = \begin{bmatrix} \sigma^2_e & \sigma_{\varepsilon,u} \\ \sigma_{u,\varepsilon} & \sigma^2_u \end{bmatrix}$$

The OLS estimator can not be applied to equation 6 because the error term not longer possess ideal properties (constant variance and zero covariance) thus the estimate would be inefficient and, hence generalized least squares (GLS) is appropriate. However, the nature of data in behavioral sciences does not permit the variance components σ^2_u and σ^2_e to be known, therefore, alternative estimation methods must be utilized. One common estimation method that can deal with the unknown variance components feasible generalized least squares (FGLS). FGLS takes an estimate of the variance components and then estimates the equation.

The individual effect in the random effect model may, too, be tested with the following hypotheses:

$$H_0: u_i = 0, \text{ or equivalently, } \sigma^2_u = 0$$
$$H_a: \sigma^2_u \neq 0.$$

After correcting the error term (ω_{it}) the *t* and *F*-test are reliable, thus inference can be regarded as valid. Based on statistical evidence, a FE model will be used in this paper.

National Education Longitudinal Data Set:88

The NELS:88 database is divided into two sections: (1) N2P, and (2) N4P. In this study, data were extracted from N2P. A representative sample of students (N=16,749) enrolled in tenth grade in the spring of 1990, who completed a questionnaire in both the first follow-up and second follow-up, were identified and used in the analysis. The LIMDEP program (Greene, 1992) and output are in the appendix.

Seven independent variables are included in the specified model. They are listed as entered into the model: (1) Review Work (F2S19BA), (2) ListenLecture (F2S19BB), (3) CopyNotes (F2S19BC), (4) Calculators (F2S19BF), (5) Think Problem (F2S20D), (6) SES (F2SES1),and (7) Self concept (F2CNCPT). The first four variables, (ReviewWork, ListenLecture, Copynotes, Calculators), are frequency measures of student educational processes and are scaled as followed: (1) Never/Rarely, (2) 1-2 Times/Month, (3) 1-2 Times/Week, (4) Almost each day, (5) Every Day. For example, the variable ReviewWork is a measure of how frequently students review their work for the previous day. The variable ListenLecture is a measure of how frequently students listen to the teacher's lecture. The variable CopyNotes is a measure of how frequently students take notes. The variable Calculator is a measure of how frequently students use calculators. The variable ThinkProblem measures student perception of teachers emphases on mathematical objectives and is scaled, (0) none, (1) minor emphasis, (2) moderate emphasis, (3) major emphasis. SES is a continuous variable

Table 1. Estimated Fixed Effects

Group	Coefficient	Standard Error	t-ratio
North East	45.84312*	0.42968	106.69
North Central	45.48305*	0.44005	103.36
South	43.06207*	0.42588	101.11
West	44.10279*	0.44472	99.17

Note. *Statistically significant at the $p < .01$.

indicating socioeconomic status. This measure is based on Duncan's (1961) socioeconomic index for all occupations. It was derived from the parent questionnaire data, the student questionnaire data, or the first follow-up or second follow-up New Student Supplement data. This variable has been standardized to have a mean of 0 and standard deviation of 1.

Selfconcept is a composite measure of all of the self-concept items (question 66) in the student questionnaire. Essentially, this variable measures students' self concept on a four point scale with: (1) strongly agree ,(2) agree ,(3) disagree,(4) strongly disagree. It should be noted that this variable was reverse scaled before a composite score was created and was standardized to have mean of zero and a standard deviation of 1. MathScore is the dependent variable and it was derived by Item response theory (IRT) to have a mean of 50 and standard deviation of 10.

Empirical Results

The FE model below was specified and estimated.

$$(Mathscore)_{it} = \beta_0 + \beta_1(ReviewWork)_{it} + \beta_2(ListenLecture)_{it} + \beta_3(CopyNotes)_{it} + \beta_4(Calculators)_{it} + \beta_5(ThinkProblem)_{it} + \beta_6(SES)_{it} + \beta_7(Selfconcept)_{it} + \varepsilon_{ty}$$

Table 1 provides evidence that the FE model is indeed the correct choice over the RE model; all t-values are significant. Region is the cross section unit (i = 1, 2, 3, 4) indicating which of the four US Census regions (1) Northeast, (2) Midwest,(3) South, or (4) West.

Table 2 provides the descriptive statistics, measure of central tendency, measure of dispersion, minimum and maximum, and number of cases. Table 3 provides the correlation coefficients for all of the variables used in the analysis.

Six out of the seven independent variables were statistically significant at the $p < 0.0001$ alpha level accounting for 25% of the variance in the dependent variable (mathematics achievement score). See Table 4 for parameter estimates.

The variable ReviewWork is statistically significant (t = 3.27, $p < .001$). As ReviewWork increases by one unit, MathScore increases by 0.234 points. In other words, as students increase the frequency in which they review their work, holding all else constant, their math score increases by 0.234 points.

Table 2. Descriptive Statistics

Variable	Mean	SD	Min	Max
MathScore (N = 12,992)	51.81	9.93	29.50	71.49
ReviewWork (N = 13,577	3.87	1.21	1.00	5.00
ListenLecture (N = 13,565)	4.24	1.03	1.00	5.00
CopyNotes (N = 13,565)	4.01	1.27	1.00	5.00
Calculators (N = 13,560)	3.69	1.38	1.00	5.00
ThinkProblem (N = 13,568)	2.23	0.84	0	3.00
SES (N = 16,563)	0.04	0.81	-3.24	2.75
Selfconcept (N = 15,123)	0.01	0.70	-3.69	1.24
Region (N = 16,426)	2.56	1.01	1.00	4.00

The variable ListenLecture is statistically significant (t = 3.924, $p < .001$). As ListenLecture increases by one unit, MathScore increases by 0.370 points. Or put differently, the more attentive the student is to the teacher's lecture, their math score increases by 0.370 points. The variable CopyNotes is not statistically significant (t = -1.488, $p < .1367$).

The variable Calculators is statistically significant (t = 11.10, $p < .001$). As Calculators increases by one unit, MathScore increases by 0.667 points. Essentially, this estimate is showing that students math score will increase with the use of a calculator.

The variable ThinkProblem is statistically significantly (t = 14.304, $p < .001$). Recall, this variable measures student perception of teachers emphasis on mathematical objectives and is scaled, (0) none, (1) minor emphasis, (2) moderate emphasis, (3) major emphasis. As ThinkProblm increases by 1 unit, MathScore increases by 1.387 point. Abstracted differently, the more teachers' emphasize "thinking about what a problem means and ways it might be solved", holding all else constant, students math score increases by 1.387 points.

The variable SES is statistically significant (t = 47.333, $p < 0.001$). As SES increases by 1 unit, holding all else constant, MathScore increases by 4.937 points. Recall SES is a continuous variable indicating member's socioeconomic status. This measure is based on Duncan's 1961 socioeconomic Index for all Occupations.

The variable SelfConcept is statistically significant (t = 10.071, $p < 0.001$). As SelfConcept

Table 3. Correlation Matrix

	Math Score	Review Work	Listen Lecture	CopyNotes	Calculators	Think Problem	SES	Self Concept
Review Work	.099	1.000						
ListenLecture	.119	.369	1.000					
CopyNotes	.103	.278	.535*	1.000				
Calculators	.148	.135	.140	.095	1.000			
Think Problem	.181	.224	.256	.237	.099	1.000		
SES	.470	.042	.044	.084	.103	.050	1.000	
Self Concept	.144	.085	.085	.076	.053	.134	.081	1.000
Region	-.103	.009	.022	-.031	.074	-.009	-.082	.002

Note. The moderate correlation between the two variables, CopyNotes and ListenLecture $r = .535$ is indicative of multicolinearity. This correlation gives reason to question the inference drawn from the *t*-ratio values on these two variables, however, the parameter estimates for CopyNotes and ListenLecture are still the best least square estimates.

increases by 1 unit, MathScore increases by 1.137 points. Essentially, students who have a more positive self perception, are scoring higher on the standardized math test.

Conclusions

In this study, a fixed effects panel data model were applied to the National Education Longitudinal Study of 1988 (NELS:88). The empirical evidence presented here suggests that student mathematics test score is influenced by educational process variables, teacher emphasis, student self-concept, and socio-economic status. Specifically, a model that included seven independent variables accounted for 25% of the variance in student mathematical achievement test score.

Caveat

The NELS:88 data set does not have a means of extracting the time component in the data. Although, models for analyzing time effects were not discussed in the study, it is an important aspect of panel data that should be coded when the data file is constructed. In addition, the time series unit should be measured in smaller periods of time. Residual analysis should be performed on the error term. That is, the error term should be analyzed for heteroscedasticity and autocorrelation.

Correspondence should be directed to:
Todd Sherron
Educational Research Lab
University of North Texas
Denton, Texas 76203-1337
E-mail: sherron@coefs.coe.unt.edu

Table 4. Fixed Effects Estimates

Variable	Coefficient	SE	*t*	*p*-value
ReviewWork	0.234	0.072	3.27	0.0011
ListenLecture	0.370	0.094	3.92	0.0001
CopyNotes	-0.190	0.073	-1.49	0.1367
Calculators	0.667	0.601	11.10	0.0000
ThinkProblem	1.388	0.097	14.30	0.0000
SES	4.938	0.103	47.73	0.0000
Selfconcept	1.137	0.113	10.07	0.0000
R^2 = 0.246		$R^2 adj$ = 0.245		

References

Green, W. (Ed.). (1997). Econometric analysis (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Greene, W. (1992). *LIMDEP User's Manual and Reference Guide*. New York: Econometric Software.

Griffiths, W., Hill, R., & Judge G., (1993). *Learning and Practicing Econometrics*. New York: Wiley.

Hasiao, C. (1985). Benefits and limitations of panel data, *Econometric Reviews*, *4*(1), 121-174.

Judge, G., Griffiths, W., Hill, R., Lütkepohl, H., & Lee, T. (1985). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.

Mátayás, L., & Sevestre, P. (Eds.). (1996). *The econometrics of panel data: A handbook of the theory with applications* (2nd ed.). Dordrecht, Netherlands: Kluwer Academic Publishers.

Tieslau, M. (1999) *Empirical Econometrics 5670*. Unpublished manuscript, University of North Texas.

Appendix

LimDep Code

```
READ;File=C:\WINDOWS\Program Files\ES Limdep\PROGRAM\nels6.lpj;
   Nobs=16749;
   Nvar=23;
   Names=x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,x15,x16,x17,x18,
      x19,x20,x21,x22,x23$

SKIP$

DSTATS; RHS = X2,X3,X4,X7,X13,X19,X20,X23; OUTPUT = 2 $

REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
       ;Str=X23
       ;Wts=X16
       ;Panel $

REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
       ;Str=X23
       ;Panel
    ;Output=2
       ;Wts=X16
       ;Fixed $
```

Data Output

```
--> SKIP$
--> DSTATS; RHS = X2,X3,X4,X7,X13,X19,X20,X23; OUTPUT = 2 $
```

Descriptive Statistics
All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
X2	3.87029535	1.21401018	1.00000000	5.00000000	13577
X3	4.24061924	1.02966250	1.00000000	5.00000000	13565
X4	4.00906612	1.27206705	1.00000000	5.00000000	13567
X7	3.69041298	1.37809060	1.00000000	5.00000000	13560
X13	2.23349057	.839799327	.000000000	3.00000000	13568
X19	.485648735E-01	.811172698	-3.24000000	2.75000000	16563
X20	.111955300E-01	.701701693	-3.69000000	1.24000000	15123
X23	2.55777426	1.01482683	1.00000000	4.00000000	16426

Correlation Matrix for Listed Variables

	X2	X3	X4	X7	X13	X19	X20	X23
X2	1.00000	.38402	.28129	.13514	.22252	.04671	.08881	.00758
X3	.38402	1.00000	.54066	.13636	.26014	.04523	.08992	.01662
X4	.28129	.54066	1.00000	.09404	.23307	.07737	.07659	-.03424
X7	.13514	.13636	.09404	1.00000	.10102	.10769	.05500	.07762
X13	.22252	.26014	.23307	.10102	1.00000	.05100	.14062	-.01871
X19	.04671	.04523	.07737	.10769	.05100	1.00000	.08907	-.08566
X20	.08881	.08992	.07659	.05500	.14062	.08907	1.00000	.01745
X23	.00758	.01662	-.03424	.07762	-.01871	-.08566	.01745	1.00000

```
--> REGRESS;Lhs=X22
       ;Rhs=X2,X3,X4,X7,X13,X19,X20
    ;Str=X23
    ;Wts=X16
    ;Panel $

+----------------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                          |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =   8, Deg.Fr.=  10887      |
| Residuals:  Sum of squares= 731652.8662    , Std.Dev.=        8.19782      |
| Fit:        R-squared=  .230712, Adjusted R-squared =          .23022      |
| Model test: F[  7,  10887] =  466.44,    Prob value =          .00000      |
| Diagnostic: Log-L = -38376.9800, Restricted(b=0) Log-L =  -39805.8044      |
|             LogAmemiyaPrCrt.=    4.208, Akaike Info. Crt.=      7.046      |
| Panel Data Analysis of X22        [ONE way]                                |
|           Unconditional ANOVA (No regressors)                              |
| Source      Variation        Deg. Free.     Mean Square                    |
| Between       20751.0              3.         6917.00                      |
| Residual      930327.          10891.         85.4216                      |
| Total         951078.          10894.         87.3029                      |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+---------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+---------+------------+
 X2        .2150986215      .72295926E-01    2.975   .0029  3.8895953
 X3        .3746630995      .95235646E-01    3.934   .0001  4.2207458
 X4       -.1516385665      .73563678E-01   -2.061   .0393  3.9716586
 X7        .7025006784      .59425942E-01   11.821   .0000  3.7239903
 X13       1.398697309      .97973819E-01   14.276   .0000  2.2173183
 X19       5.040076855      .10413200       48.401   .0000  .86701895E-01
 X20       1.035652953      .11377009        9.103   .0000  .42022957E-01
 Constant  44.54122268      .41506161      107.312   .0000

+----------------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                                   |
| Ordinary    least squares regression    Weighting variable = X16           |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189          |
| Model size: Observations =   10895, Parameters =  11, Deg.Fr.=  10884      |
| Residuals:  Sum of squares= 716924.2581    , Std.Dev.=        8.11601      |
| Fit:        R-squared=  .246198, Adjusted R-squared =          .24551      |
| Model test: F[ 10,  10884] =  355.48,    Prob value =          .00000      |
| Diagnostic: Log-L = -38266.1998, Restricted(b=0) Log-L =  -39805.8044      |
|             LogAmemiyaPrCrt.=    4.189, Akaike Info. Crt.=      7.027      |
| Estd. Autocorrelation of e(i,t)    -.000540                                |
+----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+---------+------------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+---------+------------+
 X2        .2342486655      .71596894E-01    3.272   .0011  3.8895953
 X3        .3704126069      .94402933E-01    3.924   .0001  4.2207458
 X4       -.1089542352      .73215840E-01   -1.488   .1367  3.9716586
 X7        .6668605394      .60076332E-01   11.100   .0000  3.7239903
 X13       1.387686017      .97011428E-01   14.304   .0000  2.2173183
 X19       4.937094032      .10343209       47.733   .0000  .86701895E-01
 X20       1.136968803      .11289363       10.071   .0000  .42022957E-01
```

```
+----------------------------------------------------------------------------+
|                Test Statistics for the Classical Model                     |
|                                                                            |
|        Model            Log-Likelihood    Sum of Squares    R-squared      |
| (1)  Constant term only   -39805.80420   .9510780165D+06     .0000000      |
| (2)  Group effects only   -39685.63264   .9303270109D+06     .0218184      |
| (3)  X - variables only   -38376.97986   .7316528662D+06     .2307120      |
| (4)  X and group effects  -38266.19962   .7169242581D+06     .2461983      |
|                                                                            |
|                                 Hypothesis Tests                           |
|               Likelihood Ratio Test               F Tests                  |
|           Chi-squared   d.f.  Prob.          F    num. denom. Prob value   |
| (2) vs (1)   240.343      3     .00000    80.975    3 10891     .00000    |
| (3) vs (1)  2857.649      7     .00000   466.435    7 10887     .00000    |
| (4) vs (1)  3079.209     10     .00000   355.481   10 10884     .00000    |
| (4) vs (2)  2838.866      7     .00000   462.825    7 10884     .00000    |
| (4) vs (3)   221.560      3     .00000    74.534    3 10884     .00000    |
+----------------------------------------------------------------------------+
REGR;PANEL. Could not invert VC matrix for Hausman test.

            +--------------------------------------------------+
            | Random Effects Model: v(i,t) = e(i,t) + u(i)     |
            | Estimates:  Var[e]             =   .658696D+02   |
            |             Var[u]             =   .489455D+01   |
            |             Corr[v(i,t),v(i,s)] =   .069167      |
            | Lagrange Multiplier Test vs. Model (3) = 6971.10 |
            | ( 1 df, prob value =  .000000)                   |
            | (High values of LM favor FEM/REM over CR model.) |
            | Fixed vs. Random Effects (Hausman)    =     .23  |
            | ( 7 df, prob value = 1.000000)                   |
            | (High (low) values of H favor FEM (REM).)        |
            | Reestimated using GLS coefficients:              |
            | Estimates:  Var[e]             =   .663837D+02   |
            |             Var[u]             =   .499445D+01   |
            |             Sum of Squares         .729365D+06   |
            |             R-squared              .233117D+00   |
            +--------------------------------------------------+
+---------+-------------+---------------+--------+---------+----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+---------------+--------+---------+----------+
 X2        .2341713422      .71596797E-01    3.271   .0011   3.8895953
 X3        .3703833036      .94402278E-01    3.923   .0001   4.2207458
 X4       -.1090676626      .73213889E-01   -1.490   .1363   3.9716586
 X7        .6669200976      .60069733E-01   11.102   .0000   3.7239903
 X13       1.387746607      .97011341E-01   14.305   .0000   2.2173183
 X19       4.937585087      .10343032       47.738   .0000   .86701895E-01
 X20       1.136524709      .11289247       10.067   .0000   .42022957E-01
 Constant  44.62217113      1.1804686       37.800   .0000

--> REGRESS;Lhs=X22
    ;Rhs=X2,X3,X4,X7,X13,X19,X20
    ;Str=X23
    ;Panel
       ;Output=2
    ;Wts=X16
    ;Fixed $
```

```
+-----------------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                           |
| Ordinary    least squares regression    Weighting variable = X16            |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189           |
| Model size: Observations =   10895, Parameters =   8, Deg.Fr.=  10887       |
| Residuals:  Sum of squares= 731652.8662    , Std.Dev.=       8.19782       |
| Fit:        R-squared=  .230712, Adjusted R-squared =          .23022       |
| Model test: F[  7,  10887] =  466.44,    Prob value =          .00000       |
| Diagnostic: Log-L = -38376.9800, Restricted(b=0) Log-L =  -39805.8044       |
|             LogAmemiyaPrCrt.=    4.208, Akaike Info. Crt.=      7.046       |
| Panel Data Analysis of X22        [ONE way]                                 |
|           Unconditional ANOVA (No regressors)                               |
| Source      Variation        Deg. Free.     Mean Square                     |
| Between       20751.0               3.         6917.00                      |
| Residual      930327.           10891.         85.4216                      |
| Total         951078.           10894.         87.3029                      |
+-----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+---------+-----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+---------+-----------+
 X2        .2150986215      .72295926E-01    2.975   .0029   3.8895953
 X3        .3746630995      .95235646E-01    3.934   .0001   4.2207458
 X4      -.1516385665      .73563678E-01   -2.061   .0393   3.9716586
 X7        .7025006784      .59425942E-01   11.821   .0000   3.7239903
 X13      1.398697309      .97973819E-01   14.276   .0000   2.2173183
 X19      5.040076855      .10413200       48.401   .0000   .86701895E-01
 X20      1.035652953      .11377009        9.103   .0000   .42022957E-01
 Constant  44.54122268      .41506161      107.312   .0000

+-----------------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                                    |
| Ordinary    least squares regression    Weighting variable = X16            |
| Dep. var. = X22      Mean=   52.55493863    , S.D.=   9.343603189           |
| Model size: Observations =   10895, Parameters =  11, Deg.Fr.=  10884       |
| Residuals:  Sum of squares= 716924.2581    , Std.Dev.=       8.11601       |
| Fit:        R-squared=  .246198, Adjusted R-squared =          .24551       |
| Model test: F[ 10,  10884] =  355.48,    Prob value =          .00000       |
| Diagnostic: Log-L = -38266.1998, Restricted(b=0) Log-L =  -39805.8044       |
|             LogAmemiyaPrCrt.=    4.189, Akaike Info. Crt.=      7.027       |
| Estd. Autocorrelation of e(i,t)    -.000540                                 |
+-----------------------------------------------------------------------------+
+---------+-------------+----------------+---------+---------+-----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X |
+---------+-------------+----------------+---------+---------+-----------+
 X2        .2342486655      .71596894E-01    3.272   .0011   3.8895953
 X3        .3704126069      .94402933E-01    3.924   .0001   4.2207458
 X4      -.1089542352      .73215840E-01   -1.488   .1367   3.9716586
 X7        .6668605394      .60076332E-01   11.100   .0000   3.7239903
 X13      1.387686017      .97011428E-01   14.304   .0000   2.2173183
 X19      4.937094032      .10343209       47.733   .0000   .86701895E-01
 X20      1.136968803      .11289363       10.071   .0000   .42022957E-01

        Estimated Fixed Effects
        Group       Coefficient       Standard Error       t-ratio
            1          45.84312                .42968     106.69106
            2          45.48305                .44005     103.35884
            3          43.06207                .42588     101.11422
            4          44.10279                .44472      99.16970
```

Test Statistics for the Classical Model

	Model	Log-Likelihood	Sum of Squares	R-squared
(1)	Constant term only	-39805.80420	.9510780165D+06	.0000000
(2)	Group effects only	-39685.63264	.9303270109D+06	.0218184
(3)	X - variables only	-38376.97986	.7316528662D+06	.2307120
(4)	X and group effects	-38266.19962	.7169242581D+06	.2461983

Hypothesis Tests

	Likelihood Ratio Test			F Tests			
	Chi-squared	d.f.	Prob.	F	num.	denom.	Prob value
(2) vs (1)	240.343	3	.00000	80.975	3	10891	.00000
(3) vs (1)	2857.649	7	.00000	466.435	7	10887	.00000
(4) vs (1)	3079.209	10	.00000	355.481	10	10884	.00000
(4) vs (2)	2838.866	7	.00000	462.825	7	10884	.00000
(4) vs (3)	221.560	3	.00000	74.534	3	10884	.00000

Demystifying Parametric Analyses: Illustrating Canonical Correlation Analysis as the Multivariate General Linear Model

Robin K. Henson, University of Southern Mississippi

A review of the research literature suggests that teachers need to provide students with engaging problems, facilitate their discovery of analysis methods, and encourage classroom discussion and presentation of their approaches to solving problems. The present article illustrates how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases, including multiple regression. The point is heuristic: *all* analyses are correlational, all apply weights to measured variables to create synthetic variables, and all yield effect sizes analogous to r^2. Knowledge of such relationships helps inform researcher judgement of analysis selection and use.

In one of his seminal contributions, the late Jacob "Jack" Cohen (1968) demonstrated that multiple regression subsumes all the univariate parametric methods as special cases, and thus provides a univariate general linear model (GLM) that can be employed in all univariate analyses. At about the same time, researchers increasingly also came to realize that ANOVA was being overused, and in many cases used when other methods would have been more useful. One source of ANOVA overuse was that too many researchers erroneously associated ANOVA as an *analysis* with the ability to make causal statements when using experimental research *designs*; however, it is the design, and not the analysis that leads to the ability to make definitive causal statements!

As Humphreys (1978) explained this phenomonon:

> The basic fact is that a measure of individual differences is not an independent variable [in an experimental design], and it *does not become one* by categorizing the scores and treating the categories as if they defined a variable under experimental control in a factorially designed analysis of variance. (*p*. 873, emphasis added)

Similarly, Humphreys and Fleishman (1974) noted that categorizing variables in a nonexperimental design using an ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong" (*p*. 468).

Furthermore, as Cliff (1987) noted, the practice of discarding variance on intervally-scaled predictor variables in order to perform ANOVA-type analyses creates problems in almost all cases:

> Such divisions are not infallible; think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable *more unreliable*, not less. (*p*. 130, emphasis added)

These various realizations have led to less frequent use of ANOVA methods, and to more frequent use of general linear model approaches such as regression (cf. Edgington, 1974; Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1980).

Since *all* analyses are correlational, and it is the design and not the analysis that yields the capacity to make causal inferences, the practice of converting intervally-scaled predictor variables to nominal scale so that ANOVA and other OVAs (i.e., ANCOVA, MANOVA, MANCOVA) can be conducted is inexcusable in many cases.

However, canonical correlation analysis, and not regression analysis, is the most general case of the general linear model (Baggaley, 1981; Fornell, 1978; Thompson, 1991, 1998). [Structural equation modeling (SEM) represents an even broader general linear model, but SEM is somewhat different in that this analysis usually also incorporates measurement error estimation as part of the analysis (cf. Bagozzi, Fornell, & Larcker, 1981; Fan, 1996, 1997).] In an important article, Knapp (1978) demonstrated this in some detail and concluded that "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis" (*p*. 410).

The present article will illustrate how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases. The point is not that all research ought to be conducted with canonical analyses, rather the point is heuristic: all analyses are correlational, all analyses apply weights to measured variables to create synthetic variables that become the analytic focus, and all yield effect sizes analogous to r^2 that are important to interpret. For example the R^2 obtained in a multiple regression, the eta^2 obtained from an ANOVA, and the squared canonical correlation coefficient obtained from a canonical correlation analysis all describe the variance-accounted-for between two variables and/or sets of variables. Ultimately, these statistics are directly analogous to the squared Pearson correlation.

Understanding general linear model principles aids in realizing that parametric analyses are all fundamentally related. Individual methods, such as ANOVA or *t*-tests, can then be viewed from a global perspective which will, hopefully, facilitate thoughtful researcher judgment in selecting analyses as opposed to employing "lock-step" decision strategies that limit the utility of analyses.

The Basics of Canonical Correlation Analysis

While a comprehensive discussion of CCA is beyond a scope of the present article, the reader is referred to Thompson (1991) for an accessible and user-friendly treatment of CCA. Furthermore, neither the analytic derivations of CCA nor the equivalent derivations of the linear models for the various analyses will be visited here. Since the purpose of this article is to demonstrate equivalence of models through obtained results, the reader is referred to Knapp (1978) for mathematical demonstration of the linear models.

The theory of canonical correlation analysis (CCA) has been with us for considerable time (Hotelling, 1935), but did not come into practical use until the onset of computerization (Krus, Reynolds, & Krus, 1976). In canonical analysis, the variables are considered to be members of two or more (in practice, almost always two) variable sets (e.g., pretest and posttest scores, aptitude and achievement scores) – otherwise we would analyze the data with factor analysis so as to consider simultaneously all the relationships, but without considering the existence of variable sets. Each set will include more than one variable, otherwise we generally would use a Pearson r or regression analysis. As will be shown later, these analyses are essentially the same thing anyway!

A CCA will yield many useful statistics, the most recognized of which is the *canonical correlation* (R_c). The canonical correlation describes the relationship between two synthetic variables that have been modeled from their respective variable sets by applying weights to the measured variables. A canonical correlation will be produced for each function (i.e., for each set of standardized canonical function coefficients and respective measured variables). The number of functions, each of which will be perfectly uncorrelated with the others, equals the number of variables in the smaller of the variable sets. The canonical correlation can be squared to yield a variance-accounted-for effect size (R_c^2), or the percentage of variance explainable in the criterion variable set predictable with knowledge of the variance in the predictor set.

One advantage of CCA, and other multivariate methods, lies in its *simultaneous* examination of the variables of interest, thus reducing risk of experimentwise Type I error (Fish, 1988; Henson, in press; Thompson, in press). A second, and perhaps often overlooked, advantage is the flexibility of the analysis in looking at various research problems. One example of this versatility can be found in a measurement study involving multivariate criterion-related score validity (Sexton, McLean, Boyd, Thompson, & McCormick, 1988). Thus, CCA can be used in either substantive or measurement inquiries.

Canonical Correlation Analysis as the General Linear Model

An heuristic data set for 12 elementary, middle, and high school students will be used to illustrate that CCA can conduct the other parametric methods that it subsumes, both univariate and multivariate alike. CCA will be used to perform a *t*-test, Pearson correlation, multiple regression, ANOVA, MANOVA, and descriptive discriminant analysis. Table 1 lists heuristic data on four intervally scaled variables related to motivational and personality issues: attributions of effort (EFFORT), attributions of ability (ABILIT), locus of control (LOCUS), and degree of extroversion (EXTROV). Also included are grouping data indicating some experimental treatment (TREAT) and whether students are in elementary, middle, or high school (GRADE). The reader will also notice five planned contrast variables which will be described later.

Analyses will be run using the SPSS (v9.0) statistics package. The canonical correlation macro (CANCORR) is a new addition to this version of SPSS but it limits analyses to two sets of variables of equal size. Since several examples used here include analyses on variable sets of differing size, the canonical macro was not used. There is also a General Linear Model menu which can be used to run a variety of analyses. However, for the sake of consistency and clarity, a uniform command syntax will be used in the present article to illustrate the relationship between canonical correlation the other parametric analyses. This command syntax is included in the Appendix. Note that CCA is conducted here using the MANOVA command (again, suggesting that these analyses are be related). Using Table 1 variable names, the SPSS commands for CCA are:

```
MANOVA
   LOCUS EXTROV WITH EFFORT ABILIT
   /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
   /DISCRIM=(STAN ESTIM COR ALPHA(.99)).
```

The SAS statistical software has a more direct command for CCA: PROC CANCORR. An example of SAS syntax used to perform a similar heuristic illustration can be found in Campbell and Taylor (1996).

Table 1. Heuristic Data (n=12) for Canonical Correlation Illustration

ID	EFFORT	ABILIT	LOCUS	EXTROV	GRADE	TREAT	CGR1	CGR2	CTREAT	CTGR1	CTGR2
1	10	12	18	15	1	1	-1	-1	-1	1	1
2	15	14	19	16	1	1	-1	-1	-1	1	1
3	17	18	18	13	1	2	-1	-1	1	-1	-1
4	14	13	15	10	1	2	-1	-1	1	-1	-1
5	09	15	14	04	2	1	0	2	-1	0	-2
6	06	19	16	04	2	1	0	2	-1	0	-2
7	06	20	12	07	2	2	0	2	1	0	2
8	07	19	16	03	2	2	0	2	1	0	2
9	18	11	06	18	3	1	1	-1	-1	-1	1
10	17	10	04	13	3	1	1	-1	-1	-1	1
11	12	09	10	12	3	2	1	-1	1	1	-1
12	14	13	09	14	3	2	1	-1	1	1	-1

Conducting Pearson Correlation with Canonical Correlation

When examining relationships between two variables, a Pearson correlation (r) is often invoked. The reader should immediately note conceptual similarities between a Pearson r and canonical analysis, even before examining the results from the SPSS analysis. Both investigate relationships between variables, only in the canonical case the measured variables of interest occur within multivariate sets.

A Pearson r was computed for EFFORT and ABILIT. Table 2 reports the obtained results, r = -.6150, p = .033. Table 2 also reports the CCA results, including the canonical correlation (R_c), squared canonical correlation (R_c^2), and Wilks lambda (λ). Wilks lambda, like R_c^2 is a variance-accounted-for type statistic. However, Wilks lambda indicates the variance *not* accounted for in the canonical correlation, modeled by $(1 – R_c^2)$. It is used for testing the statistical significance of R_c. As the magnitude of _ decreases (ranging from 0 to 1), the effect size (R_c^2) increases as does the likelihood of obtaining statistical significance.

For these variables, the CCA computed a squared canonical correlation coefficient of .378. The simple square root transformation of R_c^2 = .378 gives us R_c = .6148. The Pearson r and canonical correlation values are identical, save for rounding error and the fact that a canonical correlation cannot be negative. This is because the weights that are used in CCA scale the variables in the same direction, as such R_c will always range from 0 to 1. The p values are identical.

Herein lies the most fundamental of general linear model principles: all analyses are correlational. The canonical correlation is *nothing more* than a bivariate r between the synthetic variables created in CCA after the application of weights. As Thompson (1991) noted, "This conceptualization is appealing, because most researchers feel very comfortable thinking in terms of the familiar bivariate correlation coefficient" (p. 81).

Since the present heuristic CCA only had one variable in each set, the synthetic variables reflected the same relationship as did a Pearson r between the variables without the application of weights. This result should not be surprising, given the fact that multiplicative constants do not affect the value of r. The only effect the weights had in this case was to scale the variables in the same direction, thus yielding a positive value for R_c.

Conducting Multiple Regression with Canonical Correlation

As Cohen (1968) indicated, multiple regression subsumes all other univariate parametric analyses as special cases. Therefore, there is a directly analogous relationship between Pearson r and multiple regression. Since CCA subsumes Pearson r, it should be apparent that it will do the same for multiple regression.

A multiple regression analysis was conducted with EFFORT predicted by LOCUS and EXTROV. SPSS results of the regression and canonical analyses are found in Table 3. Again, all parallel statistics match within rounding error, with the exception of the weights. However, the difference between the weights is arbitrary at this point. Beta (B) weights and standardized function coefficients are easily converted into each other using the following formulas (Thompson, in press):

B / R_c = Function Coefficient

Function Coefficient * $R = B$

For example, LOCUS had a B weight of -.171156. Using R_c = .828 from the CCA, we find that the standardized function coefficient matches, within rounding error, that reported in Table 3 (-.171156 / .828 = -.2067). Since we know from the obtained results that the regression multiple R equals the canonical R_c, we can use the conversion formulas to find canonical function coefficients using only a regression analysis and B weights using only CCA.

Table 2. Conducting Pearson Correlation with Canonical (EFFORT by ABILITY)

Pearson *r* Analysis		Canonical Analysis	
r	-.615	R_c	.615
r^2	.378	R_c^2	.378
		lambda	.622
p	.033	*p*	.033

Note. R_c cannot be negative.

Table 3. Conducting Multiple Regression with Canonical (EFFORT by LOCUS and ABILIT)

Regression Analysis		Canonical Analysis	
R	.828	R_c	.828
R^2	.685	R_c^2	.685
		lambda	.315
$F(2, 9)$	9.797	$F(2, 9)$	9.797
p	.006	*p*	.006
Beta Weights		Function Coefficients	
LOCUS	-.171	LOCUS	-.207
EXTROV	.767	EXTROV	.926

Also of note here is the relationship between a Pearson *r*, the obtained multiple *R* from the regression, and the R_c from the canonical analysis. A regression analysis applies weights to observed (manifest) predictor variables to create a synthetic variable called *predicted Y* (or sometimes Y_{HAT}), which is a linear combination of the predictor variables. The multiple *R* from the regression analysis is nothing more than a Pearson correlation between predicted Y and the observed dependent measure, EFFORT in this case ($R_{\text{predicted Y, EFFORT}}$). Furthermore, as shown above, the canonical correlation (R_c) also is a Pearson *r* between two synthetic variables. In this case, however, only the predictor set (LOCUS and EXTROV) was linearly combined via the application of weights. While technically the dependent measure (EFFORT) also was transformed by a multiplicative weight, since only one variable existed, the weight was +1 and the EFFORT variable did not change. As such, the CCA and the multiple regression yielded identical results, both of which are based on a simple Pearson *r* between two variables (either manifest or synthetic)!

Conducting t-test and Point-biserial Correlation with Canonical Correlation

One of the most basic of statistical analyses is the *t*-test which is used to compare means between groups. Here a *t*-test was used to evaluate if the treatment and control groups (TREAT) differed on the EFFORT variable. Results reported in Table 4 indicate that the means of the groups were not statistically significantly different, $t = .310$, $p = .760$. A canonical analysis on the same variables yielded $F(1, 10) = .100$, $p = .760$. Note that the *p* calculated values are identical between analyses. The test statistics (*t* and *F*) are different only in metric. In fact, the *F* distribution consists of squared values of the *t* distribution. Squaring $t = .310$ produces .096 which does match the *F* value. The slight difference in the values is arbitrary and solely due to rounding error by the statistics program.

A point-biserial correlation was also conducted to illustrate the correlational nature of even the *t*-test. In essence, a *t*-test is can be conceptualized as a correlation between one dichotomous variable (TREAT) which indicates group membership and one continuous variable (EFFORT) as the dependent measure. The point-biserial correlation is a generalization of the Pearson *r* illustrated above that allows for a dichotomy in one of the variables. Again looking at Table 4, we see that the *p* values are identical across the *t*-test, point-biserial, and canonical analyses. Furthermore, the point-biserial correlation matches the magnitude of the canonical correlation within rounding error. Remember that a canonical correlation cannot be negative as discussed above. The point is again made here that all analyses are correlation in nature, even those which utilize dichotomous variables.

Conducting Factorial ANOVA with Canonical Correlation

The SPSS syntax file (see Appendix) includes commands to compute the five orthogonal contrast variables reported in the Table 1 data. Planned contrasts can be used with ANOVA methods to test specific, theory-driven hypotheses as against omnibus hypotheses (Thompson, 1994). One advantage of using planned contrasts is the ease of pinpointing statistically significant effects without having to conduct post-hoc tests which include Bonferroni-type corrections for experimentwise error. It is important to note that the contrasts will yield the same overall effect [i.e., Sum of Squares (SS) explained] as the omnibus test. They are necessary here to show that CCA can conduct ANOVA.

In the present analysis, a 3 X 2 factorial ANOVA was conducted with TREAT and GRADE as independent variables and EFFORT as the dependent variable. For the CCA, the contrast variables from Table 1 were used. The total number of orthogonal contrasts that can be created equals the degrees of freedom for each main effect. The GRADE main effect has two degrees of freedom and is represented by CGR1 and CGR2. The TREAT main effect is represented by CTREAT with one degree of freedom. CTRGR1 and CTRGR2 are simply cross products of the other main effects and test the GRADE X TREAT interaction effects. Table 5 presents results for the ANOVA: GRADE, $F = 19.367$; TREAT, $F = .510$;

Table 4. Conducting *t*-test and Point-biserial Correlation with Canonical (EFFORT by TREAT)

t-test		Canonical		Point-biserial	
$t(10)$	.314	$F(1, 10)$	.100		
p	.760	p	.760	p	.760
M(TREAT1)	12.500				
SD(TREAT1)	4.848	R_c	.100	r	-.100
M(TREAT2)	11.667	R_c^2	.010	r^2	.010
SD(TREAT2)	4.320	lambda	.990		

Note. R_c cannot be negative.

Table 5. 3 X 2 Factorial ANOVA (EFFORT by GRADE and TREAT)

Source	SS	*df*	MS	*F*	*p*	eta²
GRADE	158.167	2	79.083	19.367	.002	.743
TREAT	2.083	1	2.083	0.510	.502	.010
G x T	28.167	2	14.083	3.449	.101	.132
Error	24.500	6				
Total	212.917	11				

GRADE X TREAT, $F = 3.449$. Note that the effect size (r^2) for the error term was .1323.

Obtaining comparable results with CCA requires us to take several steps. The first step involves conducting canonical analyses in four separate designs, using EFFORT as the dependent measure and the contrasts as independent variables. Design 1 included all planned contrasts, CGR1, CGR2, CTREAT, CTRGR1, and CTRGR2, to test the total effect (SOS explained). Design 2 used CTREAT, CTRGR1, and CTRGR2 to jointly test the TREAT and interaction effects. Design 3 used CGR1, CGR2, CTRGR1, and CTRGR2 to jointly test the GRADE and interaction effects. The final CCA, Design 4, used CGR1, CGR2, and CTREAT to jointly test the GRADE and TREAT effects. Table 6 displays the Wilks' lambda values for each design from the first step. Remember that λ is something of a "reverse" effect size and will equal the effect for the error term. A quick comparison of λ for the total effect (Table 6) with the error effect size (Table 5) confirms this relationship between the statistics.

After canonical lambdas have been attained, we must use them to determine the omnibus ANOVA lambdas. This was done by dividing the Design 1 total effect (lambda) by the lambdas of the other designs. For example, to find the omnibus lambda for the GRADE main effect the total lambda (.11507) was divided by the Design 2 lambda (.85793), which reflects the joint effect of the contrast variables for the TREAT main effect and the GRADE X TREAT interaction effect. This process "removes" the effects of the other hypotheses, leaving the omnibus lambda for the GRADE main effect to be .13412516 (.11507 / .85793 = .13412516 = λ). The same process was used to find the other ANOVA lambdas with results reported in Table 6.

One final step remained. ANOVA lambdas were converted into ANOVA *F* statistics using the following formula: $[(1 - \lambda)/\lambda]*(df\text{error} / df\text{ effect}) = F$.

To illustrate, the *F* value for the GRADE main effect was modeled by [(1 - .13413) / .13413] * (6 / 2) = 19.36636. Table 6 also reports transformations for both main effects and the interaction. Note that the *F* statistics obtained by the canonical process match those obtained by the factorial ANOVA (see Table 5), within rounding error of course.

It should also be noted that the equivalence of ANOVA and CCA can be demonstrated with dummy codes that represent group membership in the independent variable (see Fan, 1978). However, the predictors would be correlated in this case. The use of orthogonal contrast is useful here to maintain the factorial structure of the groups.

Conducting Factorial MANOVA with Canonical Correlation

Since SPSS can use the MANOVA command to perform CCA, it would seem that the two are related. To illustrate the relationship, a 3 X 2 factorial MANOVA was computed with EFFORT and ABILIT as dependent variables and GRADE and TREAT as independent measures. Results from this analysis are found in Table 7. Since MANOVA is a multivariate method, Wilks lambdas are reported by SPSS and are used to test statistical significance of the *F* values.

The comparable canonical analysis was performed using the same process as with the ANOVA above. Four CCA designs using the contrast variables were run with canonical lambdas reported in Table 8. The subsequent conversion of these values to MANOVA lambdas is also found in Table 8. The reader will note the equivalence of the MANOVA _s in Table 7 with those obtained through the canonical analysis in Table 8. The final conversion to *F* values was not necessary here since the MANOVA uses the λ value to calculate *F* statistics, unlike the SOS value used in ANOVA. However, the same full model *F*-test formula used in the ANOVA section can be used to find the *F* statistics in this case.

Conducting Discriminant Analysis with Canonical Correlation

Discriminant analysis is a multivariate method that can either be used predictively to classify persons into groups or descriptively where variables identify latent structures among groups (Huberty, 1994). The descriptive discriminant analysis (DDA) case is especially useful as the preferred substitute for a one-way MANOVA or as a post hoc analysis to multi-way MANOVA analyses.

Table 6. Conduct ANOVA with Canonical Analysis (EFFORT by Contrasts)

Step One: Canonical Analyses on Four Designs			
Design	Independent Variables		lambda
1	CGR1, CGR2, CTREAT CTGR1, CTGR2		.11507
2	CTREAT, CTGR1, CTGR2		.85793
3	CGR1, CGR2, CTGR1, CTGR2		.12485
4	CGR1, CGR2, CTREAT		.24736
Step Two: Conversion of Canonical Lambdas to ANOVA Lambdas			
ANOVA Effect	Designs	Transformation	ANOVA Lambda
GRADE	1 / 2	.11507/.85793	.13412516
TREAT	1 / 3	.11507/.12485	.92166600
G x T	1 / 4	.11507/.24736	.46519243
Step Three: Conversion of ANOVA Lambdas to *F*-ratio			
Source	Transformation		*F*-ratio
GRADE	[(1-.13413)/.13413]*(6/2)		19.36636
TREAT	[(1-.92167)/.92167]*(6/1)		0.50992
G x T	[(1-.46519)/.46519]*(6/2)		3.44898

To demonstrate the DDA and CCA relationship, a descriptive discriminant analysis was conducted with TREAT as the nominally scaled predictor variable and EFFORT and ABILIT as criterion variables. Table 9 reports a statistically non-significant result $\chi^2(2, 9) = .648$, $p = .723$. The canonical analysis was conducted using the planned contrast variable CTREAT as the predictor. Results of the CCA are also reported in Table 9. The reader will note that the analyses yield identical results. One arbitrary difference is in the reporting of a χ^2 statistic for the discriminant analysis as opposed to the CCA *F* value. As with the *t* and *F* distributions described above, the difference is arbitrary since the χ^2 and *F* statistics represent the same value expressed in a different metric. The χ^2 statistic can be calculated by multiplying the *F* value by ($j * k$), where *j* is the number of variables in the predictor set and *k* is the number of variables in the criterion set. In this case, $F = .33602$, so $\chi^2 = (1 * 2).33602 = .67204$. This transformation approximates the χ^2 reported in Table 9 with the difference due to rounding.

Table 9. Conducting Multiple Regression with Canonical (EFFORT by LOCUS and ABILIT)

Discriminant Analysis		Canonical Analysis	
R_c	.264	R_c	.264
R_c^2	.070	R_c^2	.070
lambda	.931	lambda	.931
	.648	*F*	.336
df	2, 9	*df*	2, 9
p	.723	*p*	.723

Table 7. 3 X 2 Factorial ANOVA (EFFORT and ABILIT by GRADE and TREAT)

Source	lambda	*df*	*F*	*p*
GRADE	.05061	4, 10	8.61299	.003
TREAT	.61798	2, 5	1.54541	.300
G x T	.44653	4, 10	1.24122	.354

Table 8. Conduct MANOVA with Canonical Analysis (EFFORT and ABILIT by Contrasts)

Step One: Canonical Analyses on Four Designs			
Design	Independent Variables		lambda
1	CGR1, CGR2, CTREAT CTGR1, CTGR2		.03184
2	CTREAT, CTGR1, CTGR2		.62924
3	CGR1, CGR2, CTGR1, CTGR2		.05153
4	CGR1, CGR2, CTREAT		.07132
Step Two: Conversion of Canonical Lambdas to MANOVA Lambdas			
ANOVA Effect	Designs	Transformation	ANOVA Lambda
GRADE	1 / 2	.03184/.62924	.05060072
TREAT	1 / 3	.03184/.05153	.61789249
G x T	1 / 4	.03184/.07132	.44643859

Conclusion

The purpose of the present article has been to illustrate that canonical correlation analysis represents the multivariate parametric general linear model. As such, CCA can be used to conduct the univariate and multivariate analyses that CCA subsumes, including multiple regression. The point is heuristic and not intended to suggest that all analyses should be conducted with CCA. In fact, it is quite clear in the ANOVA and MANOVA examples that CCA, at least as reported by SPSS, is the long way to the same results. However, CCA would be superior to ANOVA and MANOVA when the independent variables are intervally scaled, thus eliminating the need to discard variance.

Knowing that there is a general linear model and understanding that all parametric analyses are intricately related can be of great educational value to both students and teachers of quantitative methods as well as practicing researchers. Knowing these relationships facilitates understanding of commonalties and differences among all the parametric methods and serves to inform researcher judgement concerning analysis selection and use.

The author would like to graciously thank Bruce Thompson for a review of an earlier draft of the manuscript.

Correspondence should be addressed to
Robin K. Henson
Department of Educational Leadership & Research
University of Southern Mississippi
Email: rrhenson@aol.com.

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. *Evaluation Review*, *5*, 123-131.

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*, *16*, 437-454.

Campbell, K. T., & Taylor, D. L. (1996). Canonical correlation analysis as a general linear model: A heuristic lesson for teachers and students. *Journal of Experimental Education*, *64*, 157-171.

Cliff, N. (1987). *Analyzing multivariate data.* San Diego,Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426-443.

Edgington, E. S. (1974). A new tabulation of statistical procedures used in APA journals. *American Psychologist*, *29*, 25-26.

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in *American Educational Research Journal*, *Educational Researcher*, and *Review of Educational Research* from 1978 to 1987. *Educational Researcher*, *17*(9), 19-20.

Fan, X. (1996). Canonical correlation analysis as a general analytic model. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 71-94). Greenwich, CT: JAI Press.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling*, *4*, 65-79.

Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development*, *21*, 130-137.

Fornell, C. (1978). Three approaches to canonical analysis. *Journal of the Market Research Society*, *20*, 166-181.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in *AERJ* articles, 1979-1983: The preparation of graduate students to read the educational research literature. *Educational Researcher*, *14*(2), 5-11.

Henson, R. K. (in press). Multivariate normality: What is it and how is it assessed? In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5). Stamford, CT: JAI Press.

Hotelling, H. (1935). The most predictable criterion. *Journal of Experimental Psychology*, *26*, 139-142.

Huberty, C. (1994). *Applied discriminant analysis*. New York: Wiley.

Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology*, *70*, 873-876.

Humphreys, L. G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. *Journal of Educational Psychology*, *66*, 464-472.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin*, *85*, 410-416.

Krus, D. J., Reynolds, T. S., & Krus, P. H. (1976). Rotation in canonical variate analysis. *Educational and Psychological Measurement*, *36*, 725-730.

Sexton, J. D., McLean, M., Boyd, R. D., Thompson, B., & McCormick, K. (1988). Criterion-related validity of a new standardized measure for use with infants who are handicapped. *Measurement and Evaluation in Counseling and Development*, *21*, 16-24.

Thompson, B. (in press). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics* (Vol. 2). Washington, DC: American Psychological Association.

Thompson, B. (1991). A primer on the logic and use on canonical correlation analysis. *Measurement and Evaluation in Counseling and Development*, *24*(2), 80-95.

Thompson, B. (1994). Planned versus unplanned and orthogonal versus nonorthongonal contrasts: The neo-classical perspective. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 3, pp. 3-27). Greenwich, CT: JAI Press.

Thompson (1998, April). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas*. Invited address presented at the annual meeting of the American Educational Research Association, San Diego. (Internet URL http://acs.tamu.edu/~bbt6147/)

Willson, V. L. (1980). Research techniques in *AERJ* articles: 1969 to 1978. *Educational Researcher*, *9*(6), 5-10.

APPENDIX

SPSS Command Syntax for Canonical Demonstration

```
TITLE ' Canonical correlation
        demonstration '.
TITLE ' Robin K. Henson '.
COMMENT Heuristic data for 12 cases
COMMENT EFFORT - attributions of
              effort
COMMENT ABILIT - attributions of
             ability
COMMENT LOCUS - external vs internal
              locus of control
COMMENT EXTROV - degree of
          extroversion scale
COMMENT GRADE - elementary(1),
         middle(2), high(3) school
COMMENT TREAT - treat(1),
         control(2) groups.
SET BLANKS=SYSMIS UNDEFINED=WARN
PRINTBACK LISTING.
DATA LIST
  FILE='c:\ccaasglm.txt'
  FIXED RECORDS=1
  /ID 1-2 EFFORT 4-5 ABILIT 7-8
  LOCUS 10-11 EXTROV 13-14
  GRADE 16 TREAT 18.
EXECUTE.
COMMENT Show that cca can do Pearson r.
CORRELATIONS
 /VARIABLES=EFFORT ABILIT
 /PRINT=TWOTAIL NOSIG
 /MISSING=PAIRWISE .
MANOVA
 EFFORT WITH ABILIT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
COMMENT Show that cca can do
      multiple regression.
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT EFFORT
 /METHOD=ENTER LOCUS EXTROV  .
MANOVA
 LOCUS EXTROV WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
COMMENT Show that cca can do t-test
   and point biserial correlation.
T-TEST
 GROUPS=TREAT(1 2)
 /MISSING=ANALYSIS
 /VARIABLES=EFFORT
 /CRITERIA=CIN(.95) .
MANOVA
 TREAT WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
COMMENT Show cca can do point-biserial
    which is a generalization of r.
CORRELATIONS
 /VARIABLES = treat effort
 /PRINT=TWOTAIL NOSIG
 /MISSING=PAIRWISE .
COMMENT Show that cca can do
      factorial ANOVA.
COMMENT Compute contrast variables
      to do cca.
IF (GRADE = 1) CGR1 = -1.
IF (GRADE = 2) CGR1 = 0.
IF (GRADE = 3) CGR1 = 1.
COMMENT Tests equality of the
   means of elementary(4) vs
    high school(4) students.
EXECUTE.
IF (CGR1 = -1) CGR2 = -1.
IF (CGR1 = 0) CGR2 = 2.
IF (CGR1 = 1) CGR2 = -1.
EXECUTE.
COMMENT Tests equality of means
    of middle(4) vs
    elementary high school(8) students.
IF (TREAT = 1) CTREAT = -1.
IF (TREAT = 2) CTREAT = 1.
EXECUTE.
COMMENT Tests equality of means of
   treatment (6) vs control groups (6).
COMPUTE CTRGR1 = CGR1 * CTREAT.
COMPUTE CTRGR2 = CGR2 * CTREAT.
EXECUTE.
COMMENT Tests treatment by grade
    interaction effects.
COMMENT Show contrast variables
     are orthogonal.
CORRELATIONS
 /VARIABLES=CGR1 CGR2 CTREAT
          CTRGR1 CTRGR2
  /PRINT=TWOTAIL SIG
  /MISSING=PAIRWISE .
COMMENT Step one: run factorial ANOVA
       and cca on constrast variables.
ANOVA
  VARIABLES=EFFORT
  BY GRADE(1 3) TREAT(1 2)
  /MAXORDERS ALL
  /METHOD UNIQUE
  /FORMAT LABELS .
```

```
MANOVA
 CGR1 CGR2 CTREAT CTRGR1 CTRGR2
   WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
MANOVA
 CTREAT CTRGR1 CTRGR2 WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
MANOVA
 CGR1 CGR2 CTRGR1 CTRGR2 WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
MANOVA
 CGR1 CGR2 CTREAT WITH EFFORT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
COMMENT Show cca can do MANOVA.
MANOVA
  EFFORT ABILIT BY GRADE(1 3)
        TREAT(1 2)
  /PRINT SIGNIF(MULT UNIV )
  /NOPRINT PARAM(ESTIM)
  /METHOD=UNIQUE
  /ERROR WITHIN+RESIDUAL
  /DESIGN  .
MANOVA
 CGR1 CGR2 CTREAT CTRGR1 CTRGR2
   WITH EFFORT ABILIT
 /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
 /DISCRIM=(STAN ESTIM COR).
MANOVA
  CTREAT CTRGR1 CTRGR2
    WITH EFFORT ABILIT
  /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
  /DISCRIM=(STAN ESTIM COR).
MANOVA
  CGR1 CGR2 CTRGR1 CTRGR2 WITH EFFORT
ABILIT
  /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
  /DISCRIM=(STAN ESTIM COR).
MANOVA
  CGR1 CGR2 CTREAT WITH EFFORT ABILIT
  /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
  /DISCRIM=(STAN ESTIM COR).
COMMENT Show cca can do
       discriminant analysis.
DISCRIMINANT
  /GROUPS=TREAT(1 2)
  /VARIABLES=EFFORT ABILIT
  /ANALYSIS ALL
  /PRIORS  EQUAL
  /CLASSIFY=NONMISSING POOLED .
MANOVA
  EFFORT ABILIT WITH CTREAT
  /PRINT=SIGNIF (MULTIV EIGEN DIMENR)
  /DISCRIM=(STAN ESTIM COR).
```

The Use of Problem Solving Strategies in Teaching Mathematics

Randall E. Schumacker, University of North Texas
T. Mark Beasley, St. John's University, New York

A review of the research literature suggests that teachers need to provide students with engaging problems, facilitate their discovery of analysis methods, and encourage classroom discussion and presentation of their approaches to solving problems. Two separate studies compared differences in mathematics test scores involving students randomly assigned to experimental and control conditions using a causal-comparative design. The results from both studies indicated that mathematics test scores were significantly higher for the groups of students who learned problem solving strategies. Confidence intervals, effect sizes, and bootstrap estimates are reported.

Numerous studies in mathematics education have examined the factors that are essential for learning, especially in the area of problem solving (Hudgins, 1977). For example, according to several cognitive based studies, meaningful learning is reflective, constructive, and self-regulated (Bransford & Vye, 1989; Davis & Maher, 1990; Hiebert et al., 1996; Marzano, Brandt, & Hughes, 1988; Rickard, 1995; Wittrock, 1991). Other studies have indicated that specific transfer of knowledge paradigms exist for the assessment of learning (Levine, 1975; Stolurow, 1966) and that contemporary research designs can be useful to assess transfer of learning tasks (Cormier & Hagman, 1987; Gick & Holyoak, 1987). Brooks and Dansereau (1987) have further identified four general types of learning transfer: (a) content-to-content; (b) skills-to-skills; (c) content-to-skills; and (d) skills-to-content. Snow (1989) conceptualized the learning process to include concept formation, procedural skills, learning strategies, self-regulated functions, and motivational orientations. Rosenshine, Meister, and Chapman (1996) recently reviewed numerous intervention studies and found overall that teaching students cognitive strategies for generating questions about the material improved their learning comprehension and understanding.

Identifying the important information in a problem and using that information to attempt a solution is basic to successful problem solving. Subsequent use of that information in a new problem under different circumstances presents an even higher level of problem solving skill. Problem solving, in fact, has been shown to involve at least three stages: understanding the problem, solving the problem, and answering the question (Charles, Lester, & O'Daffer, 1987; Whitener, 1989).

Palumbo (1990) further reviewed the relevant issues in problem solving research, especially the distinction between specific and generalized problem solving which focuses on the strategy required to most effectively solve a particular type of problem. Early work by Bloom and Broder (1950) has also indicated the ways in which students provide solutions to their problems: (a) gaining an understanding of the nature of the problem; (b) obtaining an understanding of the ideas contained in the problem: (c) attempting a general approach to the problem (e.g., guessing, working backwards, logical reasoning, looking for patterns); (d) using an implementation approach (no work shown, possibilities overlooked, strategy not clear); and (e) having a positive attitude and motivation toward solving the problem. Consequently, effective assessment of problem solving ability appears to require more than simply an examination of right/wrong answers given by students (Szetela & Nicol, 1992).

A further review of the literature indicated that for problem solving strategies to be effective in mathematics they must be taught (Frederikson, 1984). Rickard's (1995) case study results revealed that a teacher generally structures teaching around *their own* problem solving goals and beliefs, and not necessarily those specified in the curriculum. These findings indicated that we can not assume that a teacher has taught the necessary strategies nor allowed students the opportunity to explore and discuss their methods and solutions to a problem.

Biehler and Snowman (1990) and Ormrod (1990) have provided specific mathematics problem solving strategies that teachers can use. Their research involving 9th and 10th grade public school children enrolled in Algebra I classes indicated that students who are aware of certain problem solving strategies are more effective in working algebra problems. Overall, their findings further suggest that teachers who want students to think critically must explicitly emphasize problem solving, use varied examples, and verbalize their methods and strategies, especially if they want students to generalize, i.e., transfer, what they have learned to new and different problems.

Effective teaching, therefore, should include both the teachers' involvement in providing engaging problems and strategies using various subject matter, as well as, the teachers' facilitation of students to become more aware of *their own* metacognitive strengths and weaknesses in problem solving. Basically, in the teaching of problem solving strategies, students should be provided an opportunity to express their own strategies. The problem solving skills most commonly cited as being needed by students include: identifying the problem; distinguishing relevant from irrelevant information; choosing main points; judging the credibility of sources; making inferences from information given; observing accurately; interpreting observations; and making value judgments (National Center for Research to Improve Postsecondary Teaching and Learning, 1989-1990).

Hiebert et al. (1996) argued that reform in curriculum and instruction in mathematics should be based on allowing the student to "*problematize*" the subject, rather than mastering skills and applying them. Their method involved allowing students to contemplate why things are, to inquire, to search for solutions, resolve incongruities, and to communicate their problem solving method(s) to others. They advocated an approach based upon Dewey's "*reflective inquiry*" which involves giving engaging problems, dilemnas, and questions for the students to solve. The features of this approach are: identifying problems; active studying of the problem; and reaching a conclusion. In this context, the teachers' role is to facilitate students' analysis of the adequacy of the methods to achieve a solution to a problem. That is, the teacher should help the students to develop their own problem solving strategies.

In a recent review across several decades of research literature, Alexander (1996) addressed the role knowledge plays in learning and instruction. Findings indicated that the knowledge a learner possesses affects what information they attend to in a problem, how that information is perceived, what is judged to be relevant or important, and what is understood and remembered. One further aspect of this review suggests that a student's knowledge of topics, procedures, or strategies can be influenced by instruction. Problem solving strategies are therefore important components in the student learning process and are important factors to consider when teaching mathematics.

One could easily assume that brighter students naturally excel at problem solving in the classroom because of their high level of achievement and exemplary metacognitive ability. Related research characterizing individuals with exemplary metacognitive ability indicate they are able to: perceive large meaningful patterns; reach solutions rapidly; represent problems at a deeper level; spend more time analyzing a problem; and possess stronger self-monitoring skills (Chi, Glaser, & Farr, 1988); display their ability to learn in specific domain areas (Minsky & Papert, 1974); are better at judging the difficulty of a problem (Chi, Glaser, & Rees, 1982; Glaser, 1987); and use memory more than a general reasoning process (Posner, 1988). As a result of these findings, it seems reasonable to assume that brighter students would not benefit from instruction in problem solving strategies, however, this has never been researched. In our investigation of the use of problem solving strategies in teaching mathematics, two separate studies were conducted. We examined the effect of direct instruction of problem solving strategies on mathematics test score performance among two different groups of students: high school students and accelerated early college entrance high school students, respectively. We specifically hypothesized that students given problem solving strategy instruction would have higher average test scores on a mathematics test than students who did not receive such instruction. This approach was employed because of the research focus of our study and the overemphasis on single studies in educational research (Rosnow & Rosenthal, 1989).

We further felt that our study has significant educational importance due to the findings from the Third International Mathematics and Science Study (Beaton, et al., 1996). In 1994-1995, achievement tests in mathematics and science were administered around the world to students in classrooms. Performance expectations centered around four areas: knowing, performing routine procedures, using complex procedures, and solving problems. The United States, in comparison to other world countries, ranked among the last in mathematics test scores. This may be due more to how we teach rather than to what we teach (i.e. content).

Study One

Subjects and Design

The subjects in the first study were seventy-eight (78) 10th grade high school students who were selected for admission into an Academy of Mathematics and Science, an early college entrance program for gifted and talented students, during the spring semester. Students were accepted into the Academy based on SAT scores, personal interviews, letters of reference, and high school transcripts. The students left their respective high schools after completing 10th grade to attend the Academy full time, which was housed, on a university campus. While in the Academy, students would take university undergraduate courses in mathematics, science, and the humanities. Students who graduated from the Academy after two years concurrently received their high school diplomas and two years college credit. The average SAT-Quantitative score was 640 and the average SAT-Verbal score was

550. Students ranged in age from 15 to 18 years with 37% females, 4% African-American, 9% Hispanic, and 12% Asian-American.

Students met in a large auditorium for orientation the first week of classes at the university. After a brief presentation, students were directed to one of two different classrooms based on randomly picked seats.

The students were randomly assigned to either a control group (n=43) which *only* received the mathematics test or an experimental group (n=35) which was instructed in problem solving strategies, followed by the mathematics test. The use of an experimental control group design to investigate the effectiveness of instructional interventions has been used before (Schumacker & Miller, 1992).

Materials

A mathematics test, which included 13 problems selected from an Algebra I textbook used by high school students (Coxford & Payne, 1990), was used as the dependent measure. The types of mathematics problems selected involved skills-to-skills (e.g., arithmetic to algebra) transfer in mathematics knowledge such as (a) determining profit and loss, (b) the volume of water in different sized containers, and (c) determining the area of different shapes. Students were required to indicate their problem solving strategies for each math problem in the test booklet, i.e., methods of analysis and steps taken to answer each problem, not just provide a right/wrong answer. Each math problem had five questions worth 1 point each if correctly answered (5 points per math problem), for a maximum possible score of 65.

A standardized set of overhead transparencies was prepared which presented different strategies for the various types of mathematics problems. The problem solving strategies were adopted from Biehler & Snowman (1990) and Ormrod (1990). The strategies involved information on how various mathematical problems could be reorganized, thus leading to clues on how to solve them. Each problem on the test was different and therefore had a different problem solving strategy associated with it, so as to minimize any "teaching to the test" effect. An example math problem and problem solving strategy is in the Appendix.

Procedures

Students were directed to one of two different classrooms based upon their randomly assigned auditorium seating. One classroom represented an experimental group while the other a control group. Two different teachers were also randomly assigned to one or the other classroom. The problem solving strategies for various types of mathematics problems were presented to students in the experimental group using standardized overhead transparencies. The random assignment of the two teachers and the standardization of the materials were done to reduce any bias or teacher effects in the study. The teacher in the experimental group indicated that students had no problems or concerns about the problem solving strategy instruction provided. The two teachers and the principal author scored the mathematics tests using a scoring rubric.

Study Two

Subjects and Design

The subjects in the second study consisted of fifty-two (52) 10th grade high school students completing an Algebra I class during the fall semester. The high school students were randomly assigned to either the control group (n=25) or the experimental group (n=27). The same exact design and procedures were followed as in the first study. These students were from a different academic setting, but were of similar age and demography as those in the first study. Although the high school grade point averages were similar for students in both studies, most of the high school students in the second study had not taken the SAT. These students mainly differed from students in the first study in that they were *not* selected to attend an early college entrance program targeted for gifted and talented students.

Materials and Procedures

The mathematics test used in the first study was used in the second study. The same procedures were followed with the exception that students did not attend a university orientation session. Two high school teachers were randomly assigned to one or the other classroom. The problem solving strategies for the various types of mathematics problems were again presented to students in the experimental group using the standardized overhead transparencies. The teacher in the experimental group indicated that students had no problems or concerns about the problem solving strategy instruction provided. The two teachers and the principal author scored the mathematics tests with the same scoring rubric used in the first study.

Results

The Cronbach (1951) Alpha internal consistency reliability coefficient for the academy student scores in the first study was .84. For the high school student scores in the second study, Alpha was .85. The mean test score difference between the groups in the first study for the Academy students was 13.86. The mean test score difference between the groups in the second study for the high school students was 12.18. The test score means and standard deviations for the experimental and control groups in both studies are also presented in Table 1.

Based on the sample characteristics from each condition in these studies, O'Brien's (1981) test for unequal variances was performed (Beasley, 1995;

Ramsey, 1994). These tests verified that the experimental groups had significantly less variability in their math scores for both the Academy [$F(1,76)=9.31$, $p=.0031$] and high school students [$F(1,50)=5.14$, $p=.0028$]. Thus, it is reasonable to assume that there were differences in the variability of performance between the groups in each study.

In order to test mean differences under these circumstances, independent *t*-tests for unequal variances were performed using Satterthwaite's (1946) correction for the degrees-of-freedom (*df*). A statistically significant mean difference was found between the experimental and control group in both the first ($t=4.74$, $df=64.07$, $p=.0001$) and second ($t=3.05$, $df=36.23$, $p=.004$) studies. Thus, students in the experimental groups of both studies who received instruction in problem solving strategies had significantly higher mean test scores than students in the control groups, after the correction for unequal variances. Students in both experimental groups also demonstrated less variability in their scores, hence a need to interpret results using unequal variances. The results from both studies taken together indicate that the use of problem solving strategies in teaching mathematics is effective in improving mathematics achievement.

Post hoc Analyses

Our findings are based upon significance testing, which has recently been scrutinized because the researcher controls the sample size, level of significance, and power of the tests (e.g., Huberty, 1987; Robinson & Levin, 1997; Thompson, 1988, 1989a, 1989b, 1993; 1997). It has been recommended instead that effect sizes, confidence intervals, and bootstrap estimates be provided to better indicate the practical and meaningful interpretation of results (Kirk, 1996). Therefore, the mean differences between the groups, their respective effect sizes, and bootstrap estimates were computed and presented in Table 2.

Because mean differences were of primary interest, effect sizes were computed using a program by Mullen and Rosenthal (1985) in order to compare the results of both studies. The standard metric used for calculating the effect sizes was the standard deviation of the control group (Glass, McGaw, & Smith, 1981; Wolf, 1986). Interpretation of the effect size was based on the amount of standard deviation units the experimental group scored above the control group. It should be noted that in both studies, the control group standard deviation was previously determined to be significantly larger than the experimental group. Therefore, the mean differences reported provide conservative estimates of effect sizes.

Table 2 indicates that the Academy experimental group scored .83 standard deviation units above their respective control group, and the high school experimental group scored .69 standard deviation units above their respective control group. The gain associated with these effect sizes can be obtained by referring to a table of areas under the normal curve. Looking in a table of the areas under the normal curve, a .83 effect size corresponds to .30 of the area above the mean (above the 50th percentile). Thus, an effect size of .83 implies that if an average student in the control group were to receive instruction on problem solving strategies, they would now score at the 80th percentile of the control group. Similarly, the .69 effect size for the high school students corresponds to .25 of the area above the mean, and thus an effect size of .69 implies that if an average student in that control group were to receive instruction on problem solving strategies, they would now score at the 75th percentile of that group.

Table 1. Means and Standard Deviations of Experimental and Control Groups

Study	*n*	*Mean*	*SD*
1. *Academy*			
Control	43	43.14	16.76
Experimental	35	57.00	8.33
2. *High School*			
Control	25	45.60	17.69
Experimental	27	57.78	9.54

Bootstrap estimates and confidence intervals are also reported in Table 2 to further examine the stability of these findings. The bootstrap estimate (θ^*_B), the standard error of the bootstrap estimate, $SE(\theta^*_B)$, bias or the sampling error ($\theta^*_B - \theta$), where θ represents the contrast mean difference, and the 95% confidence interval [$\theta \pm 1.96SE(\theta^*_B)$] for each contrast were computed using programs by Lunneborg (1987). The bootstrap estimates were based upon 1,000 resampling trials.

Bias or sampling error is determined when bootstrap estimates are compared to the actual mean differences. The bias or difference between the bootstrap estimator and the sample mean differences were .20 = (14.06 - 13.86) and .10 = (12.28 - 12.18), respectively, which indicates that the magnitude of difference reflected in the means are reasonably stable estimates of the mean differences observed in the two studies (Mooney & Duval, 1993).

The 95% confidence intervals reflect the range of variation one could expect in the mean differences if conducting 1,000 replicated studies (the number of bootstrap resampling trials). The range of values for the lower and upper confidence interval estimates in both studies were similar. In the first study, the confidence intervals indicate that the mean difference between the two groups could vary between 8.12 and

Table 2. Contrasts, Effect Sizes, and Bootstrap Estimates for Experimental vs. Control Groups

Study Contrast	Effect Size[a]		Bootstrap Estimates[b]			
(1) Academy	Mean Difference	Δ	Estimator (θ^*_B)	$SE(\theta^*_B)$	Bias	95% CI
Experimental vs. Control	13.86	0.83	14.06	2.93	0.20	(8.12, 19.60)
(2) High School Experimental vs. Control	12.18	0.69	12.28	3.50	0.10	(5.32, 19.04)

Note. **a** The effect sizes (Δ) are based upon the mean difference divided by the standard deviation of the control group. (see Glass et al., 1981 for rationale on choice of metric).
b Based on 1,000 bootstrap resampling trials.

19.60. In the second study, the confidence intervals indicate that the mean difference could vary between 5.32 and 19.04.

Overall, the statistically significant mean differences, the small bootstrap estimator differences, and the 95% confidence interval values from both studies indicate strong evidence that the students in the experimental groups who were taught problem solving strategies performed better than those students in the control groups on the mathematics test.

Discussion

In two separate studies, students in the experimental group who were provided standardized instruction on problem solving strategies scored on average higher than students in a control group on a mathematics problem solving test. The first study involved 10th grade high school students who were considered above average or gifted and talented, and who had been selected to begin an early college entrance program rather than return to high school for their junior year. These students possessed high academic achievement levels and metacognitive skills, yet the experimental group of students still benefited from learning problem solving strategies. The second study involved 10th grade high school students who would be returning to complete high school. Although these students were in a different academic setting, those in the experimental group also benefited from learning problem solving strategies. The "lecture-type" presentation of problem solving strategies was practical and effective in getting the students to think about how to solve various mathematics problems and improved their mathematics test scores. The effect sizes, confidence intervals, and bootstrap estimates presented from both studies strengthen the ability to generalize the findings from these two studies to other 10th grade high school age students taking Algebra I classes.

Our findings suggest that teachers should be trained to explicitly emphasize problem solving strategies in teaching mathematics. Previous research by Biehler & Snowman (1990) and Ormrod (1990) was supported. Hiebert et al. (1996) supports the idea of a teacher using engaging problems and facilitating a students "*reflective inquiry*" so that they can discover methods to solve a problem (also see, Hiebert et al. 1997, Prawat, 1997, and Smith, 1997 for further discussion). Other research has suggested that a teacher should use a variety of examples and verbalize their methods to increase students' ability to learn and to solve problems. Hattie, Biggs, and Purdie (1996) also provide additional research support in their review of the effects of interventions on student learning. They broadly classified instructional interventions as *cognitive*, *metacognitive*, and *affective* in nature. The approach taken in this study could be characterized as a cognitive intervention because specific *tactics* were taught, which were grouped and purposefully used as *strategies* (Derry & Murphy, 1986; Snowman, 1984). Our findings should encourage teachers to address the need for using problem solving strategies during instruction (i.e., model and verbalize strategies for problem solving to their students). Given the previous research literature cited and our findings, we recommend that teachers practice giving engaging problems to students to solve, facilitate discovery of problem solving strategies and methods, use varied problem examples, and verbalize their methods and strategies, as well as, those of other students. We highly recommend that university teacher preparation programs instruct student mathematics educators in these approaches in their curriculum and instruction course work.

References

Alexander, P.A. (1996). Special issue: *The role of knowledge in learning and instruction* (Pintrich, P.R., Ed.). *Educational Psychologist*, *31*(2), 89-145.

Beasley, T.M. (1995). Comparison of general linear model approaches to testing variance heterogeneity in true and quasi-experiments. *Multiple Linear Regression Viewpoints*, *22*(1) 36-54.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996). *Mathematics Achievement in the Middle School Years*: *IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Biehler R.F., & Snowman, J. (1990). *Psychology applied to teaching*. Boston, MA: Houghton Mifflin.

Bloom, B.S. & Broder, L.J. (1950) Problem-solving processes of college students: An exploratory investigation. *Supplementary Educational Monograph*, No. *73*, Chicago: The University of Chicago Press.

Bransford, J.D., & Vye, N. (1989). *A perspective on cognitive research and its implications in instruction*. In Resnick, L.B. & Klopfer, L.E. (Eds.), *Toward the thinking curriculum: Current cognitive research*. Alexandria, VA: ASCD.

Brooks, L.W., & Dansereau, D.F. (1987). Transfer of information: An instructional perspective. In Cormier, S.M. & Hagman, J.D. (Eds.), *Transfer of learning: Contemporary research and applications*. New York: Academic Press.

Charles, R., Lester, F.K., Jr., & O'Daffer, P. (1987). *How to evaluate progress in problem solving*. Palo Alto, CA: Dale Seymour Publications.

Chi, M., Glaser, R., & Rees, F. (1982). Expertise in problem solving. *Advances in the psychology of human intelligence*, Vol. I, 17-76. Hillsdale, NJ: Lawrence Erlbaum.

Chi, M., Glaser, R., & Farr, M. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Cormier, S.M., & Hagman, J.D. (1987). *Transfer of learning: Contemporary research and applications*. New York, NY: Academic Press.

Coxford, A., & Payne, J.N. (1990). *HBJ Algebra I* (2nd Ed., p. 231). Chicago, IL: Harcourt, Brace, and Jovanovich.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Davis, R.B., & Maher, C.A. (1990). Constructivist view on the teaching of mathematics (Monograph No. 4, Chapter 1). National Council of Teachers of Mathematics, Reston, VA.

Derry, S.J. & Murphy, D.A. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research*, *56*, 1-39.

Frederikson, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research*, *54*, 363-407.

Gick, M.L., & Holyoak, K.J. (1987). The cognitive basis of knowledge transfer. In Cormier, S.M. & Hagman, J.D. (Ed.) *Transfer of learning: Contemporary research and applications*, NY: Academic Press.

Glaser, R. (1987). Thoughts on expertise. In C. Schooler & W. Schaie (Eds.), *Cognitive functioning and social structure over the life course*, pp. 81-94. Norwood, NJ: Ablex.

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: SAGE Publications, Inc.

Hattie, J., Biggs, J., and Purdue, N. (1996). Effects of learning skills intervention on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99-136.

Hiebert, J., Carpenter, T.P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, *25*(4), 12-21.

Hiebert, J., Carpenter, T.P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1997). Making mathematics problematic: A rejoinder to Prawat and Smith. *Educational Researcher*, *26*(2), 24-26.

Huberty, C.J. (1987). On statistical testing. *Educational Researcher*, *16*(8), 4-9.

Hudgins, B.B. (1977). *Learning and thinking: A primer for teachers*. Itasca, IL: Peacock Publishers.

Kirk, R.E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746-759.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing*. NJ: Erlbaum.

Lunneborg, C. (1987). *Bootstrap applications for the behavioral sciences*. User Guide: Volume 1. Department of Psychology, University of Washington. Seattle, Washington.

Marzano, R., Brandt, R., & Hughes, C.S. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: ASCD.

Minsky, M., & Papert, S. (1974). *Artificial intelligence*. Technical Reprot. Eugene, OR: State System of Higher Education.

Mooney, C.Z. & Duval, R.D. (1993). *Bootstrapping: A nonparametric approach to statistical inference* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-095). Newbury Park, CA: Sage Publications, Inc.

Mullen, B. & Rosenthal, R. (1985). *BASIC meta-analysis procedures and programs*. Hillsdale, NJ: Lawrence Erlbaum Associates.

National Center for Research to Improve Postsecondary Teaching and Learning (1989-90). *NCRIPTAL Update*, *2*(2), 1-8.

O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin, 89*, 570-574.

Ormrod, J.E. (1990). *Human learning: Theories, principles, and educational applications.* Pittsburgh, PA: MacMillan.

Palumbo, D.B. (1990). Programming language-problem solving research: A review of relevant issues. *Review of Educational Research*, *60*, 65-89.

Posner, M.I. (1988). Introduction: What is it to be an expert? In Chi, M., Glaser, R., & Farr, M. (Eds.), *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Prawat, R.S. (1997). Problematizing Dewey's views of problem solving: A reply to Hiebert et al. *Educational Researcher*, *26*(2), 19-21.

Ramsey, P.H. (1994). Testing variances in psychological and educational research. *Journal of Educational Statistics*, *19*, 23-42.

Rickard, A. (1995). Teaching with problem-oriented curricula: A case study of middle-school mathematics instruction. *Journal of Experimental Education*, *64*(1), 5-26.

Robinson, D.H. & Levin, J.R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher*, *26*(5), 21-28.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 66(2), 181-221.

Rosnow, R.L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110-114.

Schumacker, R.E., & Miller, J.E. (1992). Problem-solving differences in mathematics with accelerated students. *Texas Researcher*, *3*, 23-29.

Smith, J.P. (1997). Problems with problematizing mathematics: A reply to Hiebert et al. *Educational Researcher*, *26*(2), 22-24.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher*, *18*(9), 8-14.

Snowman, J. (1984). *Learning tactics and strategies*. In G.D. Phye & T. Andre (Eds.), *Cognitive classroom learning*: *Understanding, thinking and problem solving* (pp. 243-276). New York: Academic Press.

Stolurow, L.M. (1966). *Psychological and educational factors in transfer of training*. Media Research and Dissemination Branch, U.S. Office of Education, NDEA Title VII-A, Contract No. 4-20-002.

Szetela, W. & Nicol, C. (1992). Evaluating problem-solving in mathematics. *Educational Leadership*, *49*(8), 42-45.

Thompson, B. (1988). A note about significance testing. *Measurement and Evaluation in Counseling and Development*, *20*, 146-148.

Thompson, B. (1989a). Statistical significance result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement and Evaluation in Counseling and Development*, *2*, 2-6.

Thompson, B. (1989b). Asking "What if" Questions about significance tests. *Measurement and Evaluation in Counseling and Development*, *22*, 66-68.

Thompson, B. (1993). Special Issue: *The role of statistical significance testing in contemporary analytic practice: Alternatives with comments from journal editors*. *Journal of Experimental Education*.

Thompson, B. (1997). Rejoinder: Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher*, *26*(5), 29-32.

Whitener, E.M. (1989). A meta-analytic review of the effect on learning of the interaction between prior achievement and instructional support. *Review of Educational Research*, *59*(1), 65-86.

Wittrock, M.C. (1991). *Testing and recent research in cognition*. In M.C. Wittrock and E.L. Baker (Eds.), *Testing and Cognition*. Englewood Cliffs, NJ: Prentice Hall.

Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*(Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-059). Beverly Hills, CA: SAGE Publications, Inc.

The authors wish to express their appreciation to the MEADOWS Foundation for their external peer review and funding of this two year research project.

Correspondence should be directed to:
Randall E. Schumacker
College of Education
University of North Texas
Denton, Texas 76203-1337
E-mail: rschumacker@unt.edu

APPENDIX

Math Problem and Problem Solving Strategy

Mathematics Problem One

Assume that you have just purchased a lot on which you plan to build a home. You must tell the lender the area of your lot. Unfortunately, your lot is in the shape of a parallelogram. How do you determine the area of your parallelogram shaped lot?

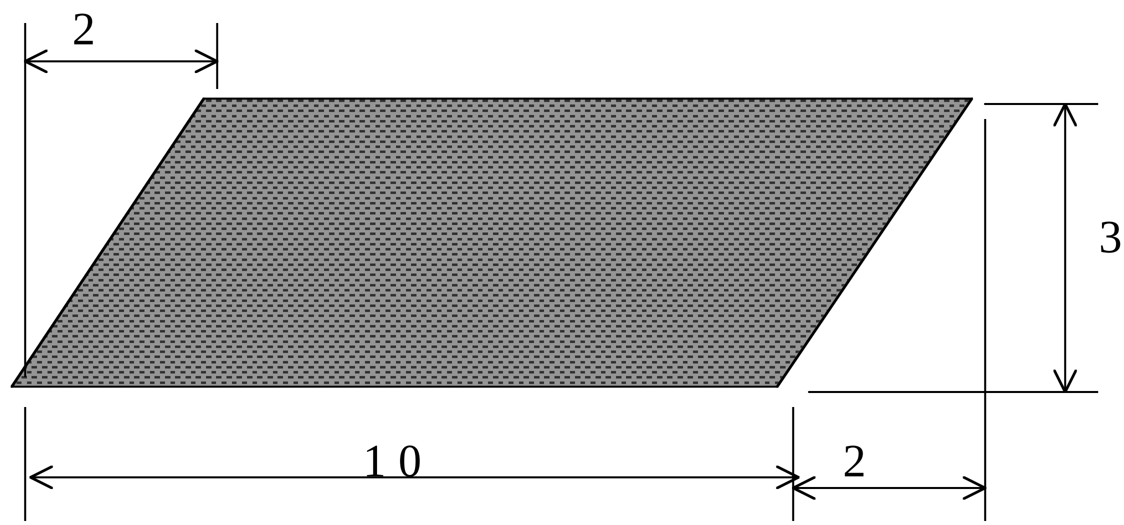
1. What is the problem?
2. What do I need to know?
3. What steps can I take to solve it?
4. What other methods could be used?
5. What is the area of your lot?

Problem Solving Strategy:

1. Convert the parallelogram into a rectangle.
2. Use the formula for determining the area of a rectangle:
   (Area = Length x Width).

Method and Solution:

1. Drop a line perpendicular to side(length).
2. Move newly formed right triangle area to opposite side
   to form a rectangle.
3. Use formula for determining area of a rectangle:
   (Area = Length x Width)



CALL FOR MANUSCRIPTS

MLRV

Multiple Linear Regression Viewpoints **needs your submissions.**

See the inside Back cover for submission details and for information on how to join the MLR: GLM SIG and get *MLRV*.

Extraneous Variables and the Interpretation of Regression Coefficients

Cam-Loi Huynh, University of Manitoba

This paper addresses some difficulties concerning the interpretation of the regression coefficients in simple and multiple regression models. The root of the problem lies in the fact that the fitted multiple regression equation is the result of transforming raw data of the independent variables into residualized scores. In the standard interpretation of the partial regression coefficients, effects of the residual term have not been explicitly differentiated from those of the regressors. Alternative interpretations of the regression coefficients are proposed. The recognition of residual and residualized effects plays an important role in the evaluation of the obtained values of the regression coefficients, R^2, the overall *F* tests and the construct validity of the multiple regression model.

There are three types of variables in a regression model, namely, the dependent variable (Y), at least one regressor or independent variable (X_j, j = 1, …, m) and the unknown error term (ε) estimated by the residual scores ($e = Y - \hat{Y}$) which in turn represent the extraneous variables, where $\hat{Y}$ is the predicted value of Y. Typically, the regression slope coefficient in the *simple* regression model $\hat{Y} = a + bX$ is defined as, "the amount of the difference in $\hat{Y}$ associated with a one-unit difference in X" (Howell, 1997, p. 242), or "The slope of the line equals the gain in Y associated with each 1-unit gain in X" (Darlington, 1990, p.10). On the other hand, each of the slope coefficients in the *multiple* regression model $\hat{Y} = a + b_1X + \ldots + b_mX_m$ is called a *partial regression coefficient* "to make clear that it is the weight to be applied to an independent variable (*IV*) when one or more specified *IV*s are also in the equation" (Cohen & Cohen, 1983, p. 83). The coefficient b_j, j = 1, 2,…, m, is defined as, "the change in the dependent variable per unit change in the jth independent variable, *assuming all other independent variables are held constant*" (Rawlings, 1988, p. 67). Similar definitions are found in several textbooks on regression analysis. It will be argued in this paper that the above definitions of b and b_j should be used with great care to avoid misleading interpretation on the effects of X_j in predicting Y for data analysis. First, some possible implications of "holding all regressors but one constant" in the multiple regression model are explored. Next, in an attempt to understand the meanings of regression coefficients, several ways to obtain their estimates are investigated. It will be demonstrated that the independent variables can be operationally transformed into the residualized terms in the process of computing the partial regression coefficients. This leads to the realization that a regression analysis transforms the obtained data into another data set called the residualized scores while reproducing the same values for the partial regression coefficients. As a result, a simple way to determine the residualized scores in multiple regression models is developed.

Before proceeding, however, an explanation of the terms "residual scores" and "residualized scores" is in order. The *residual term* (e) represents the difference between Y and $\hat{Y}$ as a result of regressing Y against one or more independent variables (X's); denoted as $e_{Y,1}$, $e_{Y,2}$, or $e_{Y,1,2}$ for regression models involving one or two regressors (where the first subscript represents the dependent variable and the subsequent subscripts, the independent variables). A *residualized variable* is formed when the residual term (e) is used either as a regressor (E_j) or as a dependent variables yielding predicted values ($\hat{e}_{Y,j}$ and $\hat{e}_{Y,h}$). The residual scores (e) capture the portion of variability in Y, called the "uncontrolled" extraneous effect of the model, that is not accounted for by *all* independent variables (X's). On the other hand, the residualized scores of X_j, say E_j (for any j = 1, …, m), represent the residual term when X_j is regressed on all other independent variables. Thus, when Y is regressed on the jth residualized variable, the resulting regression coefficient represents only the effect of X_j since the effects of other independent variables in the original multiple regression model have been "partialled out."

Winne (1969) has studied the problems of construct validity in using multiple regression models. He indicated that regressors in such models do not represent the constructs described by the original data since the partial regression coefficients are computed for the residualized scores instead. However, he did not discuss how these residualized scores can be interpreted and analyzed. Rather, Winne (1969) recommended that, "anchor variables not of direct interest in a research study be measured and correlated with residualized variables. This supplementary analysis sheds light on changes to construct validity that must be known before interpreting multiple regression analyses" (p. 187). On the contrary, it is suggested in this paper that the effects of regressors in multiple regression models are interpreted as those of residualized scores, in the same way as one would interpret partial correlation coefficients. Then, the simple regression equations of Y on the residualized variables (called the "residualized regression equations") are studied to shed light on the interpretation of the partial regression coefficients in the conventional multiple regression equation. Moreover, the coefficient of determination associated with the multiple regression model is explained in terms of the *semi-partial coefficients of determination* obtained from the above residualized regression equations. Finally, the residual plots of the multiple regression model and those of the residualized regression equations (called the *partial regression residual plots*) are examined for model diagnostics. These steps are recommended not only for identifying the effects of "controlled" extraneous variables associated with the partial regression coefficients and recapture the same R^2 but also for obtaining test statistics (t for the slopes and overall F for the fit) that take into account the influence of the residual and residualized variables. For the sake of illustration, all numerical analyses are based on the data set in Figure 1, Panel A. In the multiple regression under consideration, Test Score (Y) is regressed on Cumulative GPA (X_1) and Study Hour (X_2).

Limitations of the Conventional Interpretations

What Happens to Partial Regression Coefficients If Only Values of One Regressor Are Changed?

The main difference in the definitions of simple and partial regression coefficients given above lies in the requirement that all but one regressors in the multiple regression model are "held constant." It is true that if values of the jth regressor $X_{j,i}$ in the multiple regression equation *for the ith subject* (i = 1, 2, …, n) is changed by a constant whereas the observed values of remaining regressors are intact then the predicted value $\hat{Y}_i$ for this particular subject is modified by an amount of b_j. However, if values of X_j in the above example are modified by a fixed constant *for all subjects* then in the resulting regression equation, only the intercept term (a) will change (i.e., values of $\hat{Y}_i$ and all slopes b_1, …, b_m remain the same for i = 1, …, n). Although only values of a single regressor have been modified, one no longer has the same regression model since the intercept term has changed. The following example serves to illustrate this point.

Based on the data set in Figure 1, Panel A, three regression models are considered, the first with the original values for Y, X_1 and X_2 and the remaining two, with the linearly transformed values of $X_3 = X_1 + 5$ and $X_4 = X_2 + 3$. As expected, the resulting regression equations have the same slopes but different intercepts:

Model 1a: $\hat{Y} = a_1 + b_1X_1 + b_2X_2$
$= 43.651 + 7.301X_1 + 2.839X_2$, $R^2 = .4105$,

Model 2: $\hat{Y} = a_2 + b_1X_3 + b_2X_2$
$= 7.145 + 7.301X_3 + 2.839X_2$, $R^2 = .4105$,

Model 3: $\hat{Y} = a_3 + b_1X_3 + b_2X_4$
$= -1.373 + 7.301X_3 + 2.839X_4$, $R^2 = .4105$.

Figure 1. Data and Test Statistics for Regression Models 1a and 1b

Panel A: Data Example

ID	Y	X_1	X_2
1	73.5	3.5	2.4
2	69.0	2.8	2.5
3	85.5	3.0	5.5
4	82.0	3.7	3.1
5	90.0	3.9	5.2
6	84.0	3.1	5.5
7	86.5	3.3	7.1
8	74.5	2.9	3.6
9	71.5	3.1	5.5
10	75.5	3.6	4.4
11	80.0	4.0	5.1
12	91.8	3.5	4.2
13	86.5	3.3	7.2
Mean	81.76	3.36	1.50
SD	7.96	0.36	1.50
	$r_{Y,1} = .354$	$r_{1,2} = .354$	$r_{Y,2} = .354$

Panel B: Common Statistics for Models 1a and 1b

IV	SE	t ($p < t$)
C/Intercept	17.941	2.433 (.0322)
X_1	5.082	1.437 (.1786)
X_2	1.232	2.305 (.0416)

Panel C: Goodness-of-Fit Statistics for Models 1a and 1b

Model	SSR (SST)	R^2	MSR (MSE)	F ($p < F$)
1a	338.15 (823.77)	0.41	167.08 (44.15)	3.78 (.0547)
1b	93933.73 (94419.35)	0.99	31311.24 (44.15)	709.20 (.0001)

Note. IV = Independent variable, SE = Standard error of the regression coefficient estimate, SSR = Regression sum of squares, SST = Total sum of squares, MSR = Regression mean squares, MSE = Error mean square.

(Note that $a_2 = a_1 - 5b_1$ and $a_3 = a_1 - 5b_1 - 3b_2$.) For example, given $X_1 = 3.50$ and $X_2 = 2.40$ for the first subject then $\hat{Y} = 76.02$ and $e = 2.52$ in the three models. Typically, the same interpretation applies to the partial regression coefficients in these models, say, "If X_2 is held constant, then each 1-unit increase in X_1 leads to an average increase in $\hat{Y}$ of 7.301 units." However, if the units of measurement for any independent variable has been changed, not only a new regression model with a different intercept term is needed but also the statistical significance of the intercept term may also be altered (In the three models above, $t(\alpha) = 2.433$, $p < .03$, $t(\alpha) = 0.167$, $p > .870$

and $t(\alpha)$ = -0.032, $p > .975$, respectively). Apparently, the standard interpretation is focussed on the case in which one wants to compute a predicted value $\hat{Y}_i$, given a certain value of $X_{j,i}$, i = 1, 2, …, n, for each subject, one at a time. However, the intercept term should be considered when the scales of measurement have been changed for several, if not all, subjects. Hence, the above statement could be modified as, "For each observation i, if $X_{2,i}$ is held constant, then a 1-unit increase in $X_{1,i}$ leads to an increase in $\hat{Y}_i$ of 7.301 units. On the other hand, if all values of X_2 are changed by the same constant c, then a 1-unit increase in $X_{1,i}$ leads to an average increase in $\hat{Y}_i$ of (43.651 - 7.301c) units."

The Same Values of Partial Regression Coefficients May Not Yield the Same Regression Models

The above three regression models have different values for the intercept term but otherwise identical with respect to the test statistics of t for regression coefficients as well as overall F and R^2 for goodness of fit (as reported for Model 1a in Figure 1, Panel C). The regression coefficients in Model 1a can be reproduced by regressing Y on C, X_1 and X_2 where C is a dummy variable of constant values, say $C = 1$,

Model 1b: $\hat{Y} = a_1C + b_1X_1 + b_2X_2$

$= 43.651C + 7.301X_1 + 2.839X_2$, $R^2 = .995$.

Although the t tests for the regression coefficients in Models 1a and 1b are identical (Figure 1, Panel B), they are substantially different with respect to goodness-of-fit statistics (Figure 1, Panel C). Model 1a yields poor fit with small R^2 and marginally significant overall F. In Model 1a, R^2 represents the ratio of sum of squares of regression (SSR) over the *corrected* total sum of square (SST_c). Since Model 1b has no intercept term, R^2 has been redefined by using the *uncorrected* total sum of square (SST_u). As a result, both its R^2 and F have increased remarkably! It can be explained that this phenomenon occurs when extraneous effects independent of the predictors have been accounted for in the regression model. Hence the significance test of the failure to control for the impact of extraneous variables under the null hypothesis can be conducted by means of the following F test with degrees of freedom (q, df_r):

$$F = \frac{R^2(\text{Model.1b}) - R^2(\text{Model.1a})}{1 - R^2(\text{Model.1b})}\left[\frac{df_r}{q}\right],$$

where df_r = the residual degree of freedom in Model 1b and q = (the difference in number of regressors in Models 1b and 1a) = 1 (Darlington, 1990, pp. 124-125; Cohen and Cohen, 1983, pp. 145-151). For the data at hand, F = (.9610 - .2652)(11) = 7.653, $p <$ .00001.

Different Ways to Obtain Values of the Regression Coefficients

In an attempt to enhance the understanding, and thus improving the interpretations, of simple and partial regression coefficients, it is necessary to investigate several ways to obtain the same values of these coefficients for a given data set. Some of the steps presented below have been discussed elsewhere (Draper and Smith, pp. 196-201) but for a different objective, namely, the confirmation of the least-squares results by various methods rather than the difference in their interpretations.

As presented in Table 1, thirteen regression models can be computed on the basis of two predictors X_1 and X_2 (in Figure 1, Panel A). The three models g, h and k are the pivot models against which all remaining models will be compared. For identification purposes, the subscripts "g", "h" and "k" are attached to the regression coefficients when necessary. The predicted values $\hat{Y}$ and $\hat{X}_j$, j = 1, 2, in steps h, k, 4 and 5 are used as dependent variables ($\hat{Y}_{Y,1}$, $\hat{Y}_{Y,2}$) or regressors ($\hat{X}_1$, and $\hat{X}_2$) in steps 6 and 7, respectively. For the remaining models (steps 8 to 13), either the residual scores (obtained in steps h and k) or Y are regressed on the residualized scores (E_j obtained in steps 4 and 5) and X_j. The intercept terms are present in all regression models with raw data, except in steps 8 and 9 where only the residual and residualized scores are involved.

The regression models in Table 1 were computed using both raw and standardized data with identical variables. All the regression models with standardized scores must be fitted without the intercept terms (The computed values of the intercept terms would be zero had they been included). The results in Table 2 illustrate that it is the variable type, not the data metric, which determines the elements constituting the "extraneous variables."

Approach 1 (Based on Raw Data)

The *simple* regression coefficients for X_1 and X_2 are 7.775 (step h) and 2.911 (step k), respectively. Their *partial* counterparts are $b_{Y,1.2}$ b_{1g} = 7.301 and $b_{Y,2.1}$ b_{2g} = 2.839 (step g).

Approach 2 (Based on Standardized Scores)

For *simple* regression models (in steps h and k), the *simple*, or *zero-order*, correlations of Y and X_j are used instead of b_j (i.e., $r_{Y,1}$ r_h = .3545 and $r_{Y,2}$ r_k = .5476). For *multiple* regression models, each $r_{Y,j.i}$ denotes the *partial* correlation of Y and X_j, or the correlation of Y and X_j, given that X_i has already entered the model ($r_{Y,1.2}$ r_{1g} = .3329, and $r_{Y,2.1}$ r_{2g} = .5341 in step g).

Table 1. Regression Models for Comparing Regression/Correlation Coefficients and R^2.

Step	Regression Models
g	Y is regressed on X_1 and X_2
h	Y is regressed on X_1 (yielding $\hat{Y}_{Y,1}$ and $e_{Y,1}$)
k	Y is regressed on X_2 (yielding $\hat{Y}_{Y,2}$ and $e_{Y,2}$)
4	X_1 is regressed on X_2 (yielding $\hat{X}_1$ and E_1)
5	X_2 is regressed on X_1 (yielding $\hat{X}_2$ and E_2)
6	$\hat{Y}_{Y,1}$ (from step h) is regressed on $\hat{X}_1$ (from step 4)
7	$\hat{Y}_{Y,2}$ (from step k) is regressed on $\hat{X}_2$ (from step 5)
8	$e_{Y,1}$ (from step h) is regressed on E_1 (from step 4) (without the intercept term)
9	$e_{Y,2}$ (from step k) is regressed on E_2 (from step 5) (without the intercept term)
10	Y is regressed on E_1 (from step 4)
11	Y is regressed on E_2 (from step 5)
12	Y is regressed on X_2 and E_1 (from step 4)
13	Y is regressed on X_1 and E_2 (from step 5)

Approach 3 (Based on Predicted and Residualized Scores)

The same values of the slope/correlation coefficients in the simple and multiple regression models can also be obtained by fitting regression models on the basis of predicted ($\hat{Y}$) and residualized scores. In step 6, by regressing $\hat{Y}_{Y,1}$ obtained in step h on $\hat{X}_1$ in step 4, the *simple* regression/correlation coefficients in step h are recovered. Similarly, the results in step k are reproduced in step 7 by regressing $\hat{Y}_{Y,2}$ (step k) on $\hat{X}_2$ (step 5). The two *partial* regression/correlation coefficients in step g are reclaimed by fitting two simple regression models in terms of residualized scores ($\hat{e}_{Y,j}$ and E_j) in steps 8 and 9, respectively.

Identifying the Extraneous Variables in the Multiple Regression Model

What Are the Residualized Scores for X_j?

A much simpler procedure to obtain the residualized scores for any regressor and show that its effect can be measured by the corresponding partial regression coefficient is described below.

Step (i). Fit X_j on the remaining regressors:

$$\hat{X}_j = a + b_1X_1 + \ldots + b_{j-1}X_{j-1} + b_{j+1}X_{j+1}.$$

For Model 1a, this is realized by obtaining the regression equations in steps 4 (for X_1) and 5 (for X_2) in Table 2.

Step (ii). Obtain the residualized scores for X_j:

$$E_j = X_j - \hat{X}_j \text{ for j = 1, 2, …, m.}$$

Thus, in the example, the residual terms obtained by fitting the regression equations in steps 4 and 5 (Table 2) yield the residualized scores for X_1 and X_2, respectively.

Step (iii). Reproduce the partial regression coefficient for X_j, by fitting the regression equations in steps 8 and 9 (or steps 10 and 11, Table 2). Alternatively, they can be computed as:

$$b_j = \text{Cov}(Y, E_j)/S^2(E_j) = r(Y,E_j)\{S(Y)/S(E_j)\},$$

where E_j = the jth residualized variable, $\text{Cov}(Y, E_j)$ = covariance of Y and E_j, $S^2(E_j)$ = sample variance of E_j, $S(Y)$ = sample standard deviation of Y, and $r(Y,E_j)$ = the zero-order correlation of Y and E_j. For the example, $S(Y) = 7.9603$, $S(E_1) = .3626$, $S(E_2) = 1.496$, $r(Y,E_1) = .3325$ and $r(Y,E_2) = .5337$. Therefore, the partial regression coefficients for X_1 and X_2 are:

$b_1 = r(Y,E_1)\{S(Y)/S(E_1)\} = (.33)(7.96)/.36 = 7.30$, and $b_2 = r(Y,E_2)/\{S(Y)/S(E_2)\}=(.53)(7.96)/1.49 = 2.84$, respectively.

Understanding the Simple Regression Correlation Coefficient

So far, the regression and correlation coefficients in a simple regression model play the same roles. The same values of simple regression/correlation coefficients are reproduced in steps h and 6 (Table 2) because the predicted values of $\hat{Y}$ and $\hat{Y}_{Y,1}$ in these equations are determined by X_1 independently of X_2. Analogously, the regression/correlation coefficients in steps k and 7 are identical since the relevant predicted values of $\hat{Y}$ and $\hat{Y}_{Y,2}$ are determined by X_2 and free of X_1. Since the different values of regression and correlation coefficients are simply due to data metrics, their meanings should be interpreted similarly. The simple correlation coefficient has been defined as "a measure of the degree of closeness of the linear relationship between two variables" (Snedecor and Cochran, 1967, p. 173). This statement remains meaningful in the context of simple regression models with either raw or standardized scores. A linear relationship is one in which the variation in Y, produced by a specified change in X, is constant. The "linear relationship" between X and $\hat{Y}$ in the simple regression model with the intercept has two components, constant (determined by the intercept) and linearly changeable (accounted for by the slope). Therefore, the slope regression coefficient in a simple regression model can be interpreted as, "*In raw data metric, the slope b represents the relative weight of X to account for the linear variability in $\hat{Y}$ that is free of the unknown extraneous effects represented by the residual $e = Y - \hat{Y}$.*"

In other words, bX represents the linear trend of, or portion of the linear variation in, the values of $\hat{Y}$ that is not attributed to unknown extraneous effects.

Table 2. Results and Test Statistics for the Regression Models in Table 1.

Step	Regression Model (Raw Data)	Regression Model (Standardized)	SSR (SST)	R^2	MSR	MSE	F ($p<F$)
g	$\hat{Y} = a_g + b_{1g}X_1 + b_{2g}X_2$ $= 43.652 + 7.301X_1 + 2.839X_2$	$\hat{Y} = r_{1g}X_1 + r_{2g}X_2$ $= .3329X_1 + .5341X_2$	338.15 (823.77)	.4105	169.07	44.15	3.83 (.05)
h	$\hat{Y} = a_h + b_hX_1$ $= 55.606 + 7.775X_1$	$\hat{Y} = r_hX_1 = .3545X_1$	103.50 (823.77)	.1256	103.50	60.06	1.72 (.00)
k	$\hat{Y} = a_k + b_kX_2$ $= 67.874 + 2.911X_2$	$\hat{Y} = r_kX_2 = .5476X_2$	247.03 (823.77)	.2999	247.03	48.06	5.14 (.04)
4	$\hat{X}_1 = a_1 + b_{1,2}X_2$ $= 3.317 + .010X_2$	$\hat{X}_2 = r_{2,1}X_1$ $= .0404X_1$	0 (1.71)	.0016	0	0.14	0.02 (.89)
5	$\hat{X}_2 = a_2 + b_{2,1}X_1$ $= 4.210 + .167X_1$	$\hat{X}_1 = r_{1,2}X_2$ $= .0404X_2$	0.05 (29.15)	.0016	0.04	2.42	0.02 (.89)
6	$\hat{Y}_{Y,1} = a_h + b_h\hat{X}_1$ $= 55.606 + 7.775\hat{X}_1$	$\hat{Y}_{Y,1} = r_{h}\hat{X}_1$ $= .3545\hat{X}_1$	0.17 (103.50)	.0016	0.17	8.61	0.02 (.89)
7	$\hat{Y}_{Y,2} = a_k + b_k\hat{X}_2$ $= 67.874 + 2.911\hat{X}_2$	$\hat{Y}_{Y,2} = r_{k}\hat{X}_2$ $= .5476\hat{X}_2$	0.40 (247.03)	.0016	0.40	20.55	0.02 (.89)
8	$\hat{e}_{Y,1} = b_{1g}E_1$ $= 7.301E_1$	$\hat{e}_{Y,1} = r_{1g}E_1$ $= .3329E_1$	234.65 (720.27)	.1580	234.65	37.36	6.28 (.03)
9	$\hat{e}_{Y,2} = b_{2g}E_2$ $= 2.839E_2$	$\hat{e}_{Y,2} = r_{2g}E_2$ $= .5341E_2$	91.12 (576.74)	.3258	91.12	37.36	2.44 (.14)
10	$\hat{Y} = \bar{Y} + b_{1g}E_1$ $= 81.764 + 7.301E_1$	$\hat{Y} = r_{1g}E_1$ $= .3329E_1$	91.12 (823.77)	.1106	91.12	61.05	1.49 (.25)
11	$\hat{Y} = \bar{Y} + b_{2g}E_2$ $= 81.764 + 2.839E_2$	$\hat{Y} = r_{2g}E_2$ $= .5341E_2$	234.65 (823.77)	.2840	234.65	49.09	4.78 (.05)
12	$\hat{Y} = a_k + b_kX_2 + b_{1g}E_1$ $= 67.874 + 2.911X_2 + 7.301E_1$	$\hat{Y} = r_kX_2 + r_{1g}E_1$ $= .5476X_2 + .3329E_1$	338.15 (823.77)	.4105	169.07	44.15	3.83 (.05)
13	$\hat{Y} = a_h + b_hX_1 + b_{2g}E_2$ $= 55.606 + 7.775X_1 + 2.839E_2$	$\hat{Y} = r_hX_1 + r_{2g}E_2$ $= .3545X_1 + .5341E_2$	338.15 (823.77)	.4105	169.07	44.15	3.83 (.05)

Note. For variables with double subscripts, the first subscript refers to the dependent variable and the second subscript denotes the regressor.

Moreover, $a + bX$ constitutes the value of $\hat{Y}$ with the maximum value of R^2 if it can be assumed that the influence of the unknown extraneous effects is equally distributed to all members of the sample. This assumption can be checked by running the regression model without the intercept that contains X and a dummy variable C of fixed values. As a result, the same value for b in the original simple regression equation is reproduced in the multiple regression model without the intercept. For the regression models in steps h and k, by letting $C = 1$ for all subjects, say, we get

$$\hat{Y} = a_hC + b_hX_1 = 55.606C + 7.775X_1,\quad R^2 = .9954, \text{ (revised step h)}$$

$$\hat{Y} = a_kC + b_kX_2 = 67.874C + 2.911X_1,\quad R^2 = .9939, \text{ (revised step k)}$$

The values of R^2 have been increased dramatically (as compared to those reported for steps h and k in Table 2) to reflect the fact that extraneous effects have been (artificially or statistically) controlled.

Understanding the Partial Regression Correlation Coefficient

The residual term in step h is regressed on the residual term in step 4 to produce the residualized scores $\hat{e}_{Y,1}$ in step 8, representing the portion of variation in Y that is free of X_2. Analogously, the residual terms in steps k and 5 are used to yield the residualized scores $\hat{e}_{Y,2}$ in step 9 representing the part of variation in Y that is not influenced by X_1. Therefore, b_{1g} (in steps g and 8) denotes the *relative weight* of X_1 in the raw data metric (or r_{1g}, the *simple correlation* between Y and X_1 in terms of standardized scores) *that is free of* X_2. Analogously, b_{2g} (in steps g and 9) signifies the *relative weight* of X_2 in raw data metric (or r_{2g}, the *simple correlation* between Y and X_2 based on standardized scores) *that is free of* X_1. Since the partial regression coefficients in step g can be reclaimed as two simple regression coefficients in steps 8 and 9, the simple and partial regression coefficients should be logically defined and explained similarly. This is the approach adopted in the following discussion.

The partial correlation coefficient, say $r_{1,2.3}$, is commonly defined as, "the correlation between variables 1 and 2 in a cross section in individuals *all having the same value* of variable 3" (Snedecor and Cochran, 1967, p. 400). In the regression context, the partial correlation $r_{Y,1\ .\ 2,\ ...,\ m}$, say, represents the portion of the correlation of Y and X_1 which has no dependence on values of the variables X_2, ..., X_m and extraneous effects. In the same vein of logic, the partial regression coefficients for X_j in a multiple regression model can be interpreted as, "*In raw data metric, the slope b_j represents the relative weight of X_j to account for the linear variability in $\hat{Y}$ that is free of the effects due to other regressors in the model and the unknown extraneous effects*."

The effects due to other regressors are estimated by the residualized scores $\hat{e}_{Y,j}$ (steps 8 and 9) whereas the unknown extraneous effects are estimated by the residual $\hat{Y}$ - Y (step g). The meaning of this interpretation is further explained by the two multiple regression equations in steps 12 and 13. The partial regression coefficient b_k in step 12 represents the simple regression coefficient (or relative weight) of X_2 whereas b_{1g}, the partial regression coefficient in step g, is actually transformed into an extraneous effect, being the slope of a residualized variable (E_1). A similar interpretation applies to b_h and b_{2g} in step 13. The transformation of regressors into residual and residualized variables in the multiple regression models is not expected to influence the test statistics. Indeed, as shown in Table 2, R^2, the sum of squares, mean squares, and F of the three models g, 12 and 13 are identical.

As shown above, had the extraneous effects been "controlled" by a dummy variable, say C = 1, the regression model in step g can be reproduced but with a much greater value of R^2.

Implications of Taking Extraneous Variables into Consideration

There are at least three pertinent outcomes rendered by the recognition of extraneous effects in the regression model: (i) an understanding of the limitations in the construct validity of multiple regression models, (ii) a proper decomposition for the coefficient of determination (R^2), and (iii) an improvement in the evaluation of estimates of the regression/correlation coefficients and the overall F tests.

Construct Validity in Multiple Regression Analysis

The extent to which the regressors can be used to meaningfully explain and accurately predict values of the dependent variable represents the construct validity of the regression model. The analysis so far indicates that, although the regressors (X_1, ..., X_m) are used in the multiple regressions, the regression slopes and R^2 measure the contributions of the residualized scores (E_1, ..., E_m) or a mixture between regressors and residualized scores unless X_1, ..., X_m are uncorrelated. Hence, the construct validity in multiple regression analysis may be low. The following illustration is adapted from Winne (1989), given the results in Table 2. From the three basic regression models (in steps g, h and k), how do the relationships among Y, X_1 and X_2 be explained? One may be tempted to arrive at the following conclusions:

(i) If the entry order was X_1 and X_2 then X_1 accounted for 12.56% of the variability in Y and X_2 accounted for an additional 28.49% of the variability in Y (since R^2 = .4105 in step g, R^2 = .1256 in step h and .4105 - .1256 = .2849). On the other hand, if the entry order was X_2 and X_1 then X_2 accounted for 29.99% of the variability in Y and X_1 accounted for the remaining 11.06% of the variance in Y.

(ii) When all variables are transformed to standardized scores, an increment of one standard deviation in X_1 is associated with a 33.29% increase in Y. Similarly, an increment of one standard deviation in X_2 yields an increase in Y by 53.41 percent.

Although intuitively meaningful, both of these statements are wrong with respect to the revised interpretations of partial regression/correlation coefficients! In the first statement (i), for the (X_1, X_2)-entry order, X_2 did not account for the additional 28.49% of the variability in Y but the residualized scores E_2 did. The statement is correct if X_2 is replaced by E_2. This can be seen by following the series of equations in steps h, 11 and 13 (either raw data or standardized scores). The last model (step 13) contains the same values for the slopes and the sum of R^2's reported for the combination of models h and 11. Similar arguments apply to the (X_2, X_1)-entry order in the second part of statement (i) above based on the results for steps k, 10 and 12. Statement (ii) is wrong since X_1 and X_2 are correlated. The statement is correct by either of the following modifications. First, the percentages are changed to 35.45% and 54.76% for X_1 and X_2, respectively (see steps h and k). The pairs (33.29%, 53.41%) and (35.45%, 54.76%) are quite close to each other since the correlation between X_1 and X_2 is quite small ($r_{1,2}$ = .04). Greater difference is expected for larger $r_{1,2}$. Alternatively, X_1 and X_2 are replaced by E_1 and E_2, respectively (see steps 10 and 11).

The mistakes made in statements (i) and (ii) presage a serious error that materializes when one attempts to assess the statistical significance of the slopes and determine the proportional contributions of the regressors to variations in Y in multiple regression models. For these purposes, the results of Table 2 should be obtained and examined in conducting the statistical evaluation of the standard multiple regression model. In particular, the regression models in steps 10 and 11 in terms of residualized scores should be used for studying the statistical inference of the partial regression/correlation coefficients. We return to this point later. Meanwhile, the following discussion serves to illustrate how to analyze the multiple regression model taking into consideration the effects of residualized scores and extraneous factors.

Decomposition of the Coefficient of Determination

The decomposition of R^2 for the multiple regression model is given by Engelhart (1936) as, say for m = 3,

$$R^2 = \beta_{Y,1}{}^2 + \beta_{Y,2}{}^2 + \beta_{Y,3}{}^2 + 2\beta_{Y,1}\beta_{Y,2}r_{1,2} + 2\beta_{Y,1}\beta_{Y,3}r_{1,3} + 2\beta_{Y,2}\beta_{Y,3}r_{2,3}, \quad (1)$$

where $\_{Y,j}$ = the standardized partial regression coefficient of X_j and $r_{j,h}$ = the zero-order correlation of X_j and X_h. From this equation, it was argued that the total variance in Y is reproduced by the direct variance (indicated by the betas squared) and shared variance (denoted by twice the sum of the correlational cross products) of the regressors. Engelhart (1936) argued that the shared variance is divided among each of the regressors in the same proportions as the direct variance. Chase (1960) modified this equation to be

$$R^2 = (\beta_{Y,1}{}^2 + \beta_{Y,1}\beta_{Y,2}r_{1,2} + \beta_{Y,1}\beta_{Y,3}r_{1,3}) + (\beta_{Y,2}{}^2 + \beta_{Y,1}\beta_{Y,2}r_{1,2} + \beta_{Y,2}\beta_{Y,3}r_{2,3}) + (\beta_{Y,3}{}^2 + \beta_{Y,1}\beta_{Y,3}r_{1,3} + \beta_{Y,2}\beta_{Y,3}r_{2,3}). \quad (2)$$

so that “the total direct and shared variance in the criterion associated with the ith independent variable is given by the square of the beta for the ith variable, plus half of all the covariance terms in formula (1) which include the beta for the ith variable” (p. 266). The decomposition of R^2 for step g in Table 2 yields:

Direct effect of X_1: $\beta_{Y,1}{}^2 = (.3329)^2 = .11082$,
Direct effect of X_2: $\beta_{Y,2}{}^2 = (.5341)^2 = .28526$,
Shared effect of X_1 and X_2:
[$\beta_{Y,1}\beta_{Y,2}r_{1,2}$= (.3329)(.5341)(.04044) = .00719],
Total effect of X_1: .11082 + .00719 = .11802,
Total effect of X_2: .28526 + .00719 = .29245,
The multiple coefficient of determination:
R^2 = .11802 + .29245 = .41047.

Although this decomposition reproduces the multiple coefficient of determination (R^2), it is not useful in determining the contribution of the residualized variables since the coefficients of determination in steps 10 and 11 (Table 2) are not equal to the total effects of X_1 and X_2, assumed in (2) as the components of R^2. Moreover, the decomposition (2) "has none of the most important properties that a "contribution to variance" has when variables are uncorrelated" (Darlington, 1968, p. 170). A more appropriate partition of R^2 is based on the semi-partial coefficients of determination. The general form of the semi-partial coefficient of determination for the jth residualized variable ($R^2_{Y,j.j*}$) is $R^2_{Y,j.j*} = R^2 - r^2_{Y,j}$, for j ≠ j* = 1, …, m, where R^2 = the (multiple) coefficient of determination of the full model and r_{Yj} = the zero-order correlation of Y and X_j. The semi-partial coefficients of determination for X_1 and X_2 in the example are $R^2_{Y,1.2} = R^2 - r^2_{Y,1} = .4105 - (.3545)^2 = .2849$, and $R^2_{Y,2.1} = R^2 - r^2_{Y,2} = .4105 - (.5476)^2 = .1106$, respectively. As a result, the coefficient of determination in the multiple regression model can be expressed as $R^2 = \{R^2_{Y,1.2} + R^2_{Y,2.1} + r^2_{Y,1} + r^2_{Y,2}\}/2$.

Effects of Extraneous Variables on Statistical Inference

In analyzing the goodness of fit of the multiple regression model, the researcher would get a clearer understanding of the role played by partial regression coefficients by fitting the conventional (step g) and residualized versions (steps 10 and 11). The model in step g has the advantage that the regressors are expressed in terms of the original unit of measurement. Hence, with a reasonable R^2, it can be used for predicting Y. However, in assessing the contributions of the regressors to variations in Y, the regression coefficients of the residualized scores in steps 10 and 11 are more meaningful and should be used.

For the multiple regression model in step g, the slope of X_2 is statistically significant at $\alpha = .05$ [$t(b_2) = 2.305$, $p < .05$] whereas that of X_1 is not [$t(b_1) = 1.437$, $p > .18$]. However, the significance of X_2 may be misleading in light of the overall F statistic ($p > .05$, Table 2). On the other hand, the regression models of Y using the residualized variables in steps 10 and 11 facilitate the evaluation of the statistical inference on the regressors in the multiple regression model (step g). Evidently, both E_1 and E_2 are not statistically significant ($p > .25$ and .05, respectively in Table 2). Whereas the multi-dimensional graph of Y against X's that also contains the regression line for the multiple model in step g is hard to draw, the regression lines of the residualized variables can be easily depicted since the simple models 10 and 11 involve only single regressors (E_1 or E_2) and their intercept term is equal to the sample mean of Y. The plot of the regression line for step 10, say, is the same as the plot of Y on X_1 at given values of X_2 (as illustrated by Mullet, 1972) but with much less effort.

Conclusions

It is suggested that the partial regression coefficient b_j represent the effect of the jth residualized variable which is computed as the difference between X_j and its predicted values obtained by regressing X_j on all other independent variables in the multiple regression model. The revised interpretations of the regression coefficients are based not only on the mathematical properties of the regression equation but also on the sources of the values reported by such an equation. The proposed interpretations of simple and partial regression coefficients reflect the same meanings conveyed by their corresponding correlation coefficients. The consideration of residualized effects in regression analysis leads to explanations that are more uniform in terminologies for both simple and partial regression coefficients. Moreover, it enables a recognition of low construct validity in regression modelling and sheds light on how to analyze the test statistics in fitting regression models. In the simple regression model, due to the lack of residualized variables, the simple regression coefficient for X_j measures its effect in predicting Y without recognizing extraneous variables. On the other hand, the partial regression coefficient for X_j measures its contribution in predicting Y when the extraneous effects to X_j generated by *all other regressors* have been explicitly accounted for. In all regression models, the remaining effects of the extraneous variables are represented by the residual term (e).

References

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin, 69*, 161-182.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Howell, D. C. (1997). *Statistical methods for Psychology* (4th ed.). Belmont, CA: Duxbury.

Mullet, G. M. (1972). A graphical illustration of simple (total) and partial regression. *The American Statistician, 26*, 25-27.

Rawlings, J. O. (1988). *Applied regression analysis. A research tool*. Pacific Grove, CA: Brooks/Cole.

Snedecor, G. W. & Cochran W. G. (1967). *Statistical methods* (6th ed.). Ames, IA: The Iowa State University Press.

Winne, P. H. (1983). Distortions of construct validity in multiple regression analysis. *Canadian Journal of Behavioral Science, 15(3)*, 187-202.

This research was partially supported by a grant from Social Sciences and Humanities Research Council of Canada (SSHRC No. 410-98-1239).

Correspondence should be addressed to
Cam-Loi Huyhn
Department of Psychology
University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2
Email: Huynh@cc.umanitoba.edu.

Path Model of Treatment Outcome in a Multidisciplinary Pain Management Clinic

Daisha J. Cipher, Southern Methodist University
P. Andrew Clifford, Cognitive Psychophysiological Institute at Dallas

The treatment of chronic pain disorders has become multifaceted as the field of pain research has recognized the complex nature of chronic pain. Multidisciplinary pain management has been developed in order to address the complexities of chronic pain disorders. However, in the study of multidisciplinary pain management, there have been few models predicting patients' response to treatment. This study examined a path model of treatment outcome, incorporating such variables as coping styles, treatment compliance, and treatment outcome. Results indicated that a coping style involving the suppression of negative emotion is associated with more treatment compliance, functional capacity, and perceived life control. A coping style involving amplification of negative emotion was found to be associated with poorer treatment compliance, functional impairment and emotional distress, such as depression and anxiety. Possessing an aggressive coping style was found to be associated with poor treatment compliance, as well as anger, hostility, and a low probability of benefiting from a treatment program.

Chronic pain is reported by 80 million Americans (Bonica, 1987), and 60% of all social security disability claims involve the allegation of pain (Simmons, Avant, Demski, & Parisher, 1988). In view of the vast empirical support for psychological treatments for pain, the strict biomedical intervention for pain has given way to multidisciplinary pain management (Flor, Fydrick, & Turk,. 1992). Multidisciplinary pain management typically incorporates not only pharmacotherapy and physical therapy, but also biofeedback, operant conditioning, relaxation and cognitive restructuring. The most common goals of multidisciplinary pain centers (MPCs) are functional capacity, pain reduction, reduction in addictive medication, reduction of health-care utilization, increased activity including return to work, closure of disability claims, and reduction in emotional distress, with functional capacity considered most important, by clinicians and insurance companies alike (Turk, 1996).

In the MPC treatment outcome research, most salient is the need for predictors of success (and for that matter, failure) of MPC treatment. Such predictors would allow clinicians to identify those patients who will benefit from an MPC approach, and those who might need an alternative form of treatment. Preliminary studies have indicated that the coping scales of the Millon Behavioral Health Inventory (MBHI) are good predictors of behavioral treatment outcomes (Wilcoxson et al., 1988, Gatchel et al., 1985), and there is evidence that these scales can be used to classify distinct coping styles of chronic pain patients (Dickson et al, 1992, Cipher & Clifford, 1996; Marron et al, 1984). The MBHI coping scales are described below (see Table 1).

A recent factor analysis (Cipher, 1999) performed on the eight MBHI coping styles confirmed past cluster analytic findings of Cipher & Clifford (1996), as well as the actual authors of the MBHI. The factors extracted are summarized below:

Factor One: Expression of Negative Emotion. The MBHI Inhibited and Sensitive Scales loaded negatively on Factor One, consistent with other studies finding these scales to be grouped together (Dickson et al., 1992; Gatchel et al, 1985; Marron et al., 1984). The Confident and Sociable scales loaded positively on Factor One. Based on these loadings and the results from the correlational analyses, Factor One appears to be a dimension of expression of negative emotion. That is, on one end of the dimension, there appears to be a high reporting of emotional distress and neuroticism. On the other end, there is an underreporting of distress coupled with high defensiveness. For example, Factor One is negatively correlated with affective distress, functional impairment, depression, and overall psychopathology (MPI I, MPI AD, MMPI-2 D, MMPI-2 F scales). Factor One is positively correlated with a subjective sense of life control (MPI LC scale), and positively correlated with defensiveness and the denial of psychopathology (MMPI-2 K and F scales, respectively) – similar to a "Polyannish" attitude. Thus, on one end of the dimension, there is suppression of negative emotion, and on the other end, "amplification" of negative emotion. Consequently, it appears that Factor One has largely captured the clusters found by Cipher and Clifford (1996) onto one dimension, with suppression of negative emotion and stress on one end (e.g. Repressors), and amplification of negative emotion on the other (e.g. Amplifiers).

Factor Two: Aggression. The Cooperative scale loaded negatively on Factor Two, while the Forceful scale loaded positively. This factor appears to be a dimension of aggression. One end of the dimension represents aggression and forcefulness. The other end represents passiveness and cooperation. Correlational analyses revealed Factor Two to be positively related to anger, cynicism, anti-social practices and Type A behavior (MMPI ANG, CYN, ASP, and TPA scales, respectively). Factor Two is not related to neuroticism per se; rather, it is associated with anger, hostility, resentment of authority, having a temper, being impatient, and being critical. Factor Two was negatively related to defensiveness, and is associated with frankness and self-centeredness (MMPI K scale). In sum, Factor Two appears to be separate from amplification; it is a dimension of active independence, anger, and resentment on one end, and passive dependence and cooperation on the other.

This study examined the role of coping styles in the chronic pain patient's treatment compliance and outcome in order to identify those patients who respond (and do not respond) to multidisciplinary pain management. Patients' compliance with their treatment regimen is an important factor in any clinical setting, but is often overlooked when examining treatment outcome and cost effectiveness. In one of the few studies quantifying treatment compliance in the pain management context, Lutz, Silbret and Olshan (1983) found a significant relationship between compliance and treatment outcome. However, compliance has not been examined as a mediator between coping/personality styles and outcome.

The findings of Cipher and Clifford (1996) indicated that certain coping styles might be predictive of chronic pain patients' treatment compliance and post treatment outcome. As outlined by Turk (1996), functional impairment is one of the most common and useful outcome variables examined in multidisciplinary pain centers. This study assessed the predictive value of the MBHI coping styles in a cognitive-behavioral pain management treatment outcome model, with treatment compliance as a mediator between coping styles and treatment outcome. Figure 1 below illustrates the proposed model of treatment outcome.

Table 1. Brief Descriptions of High Scorers on the MBHI Coping Style Scales

Style	Description
Introversive	Keeps to self, quiet, unemotional, not easily excited, lacks energy
Inhibited	Shy; socially ill-at-ease, avoids close relationships, fears rejection
Cooperative	Soft-hearted, reluctant to assert self, submissive, dependent
Sociable	Charming, emotionally expressive, histrionic, talkative
Confident	Self-centered, egocentric, acts self-assured
Forceful	Domineering, abrasive, intimidates others, blunt, aggressive
Respectful	Serious-minded, efficient, rule conscious, emotionally constrained
Sensitive	Unpredictable, moody, passively aggressive, negativistic

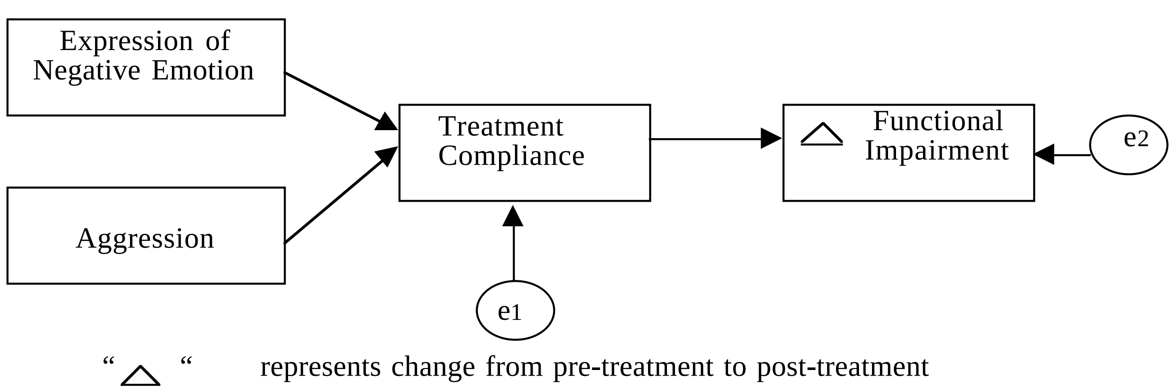
Method

Data were collected from 67 outpatients who completed treatment at a University pain clinic. All patients had been previously diagnosed with some sort of chronic pain syndrome. Exclusion criteria were the presence of any cognitive deficits due to neurological disorders, progressive terminal illnesses, or any other medical conditions which were not stable (e.g. end-stage cancer). The most common diagnoses were low back pain, neck/shoulder pain, headache, neuropathy, and fibromyalgia. The average age of patients was 45 years old.

Treatment.

Treatment consisted of multi-disciplinary pain management, which included cognitive-behavioral therapy incorporating biofeedback and relaxation train-

Figure 1. Path Model of MPC Treatment Outcome



ing, and pharmacotherapy. Licensed psychologists provided cognitive-behavioral therapy. Pharmacotherapy was provided on a monthly basis by attending anesthesiologists. Goals of pharmacotherapy involved tapering the patients to lowest possible dosages of analgesics required to minimize the pain.

Measures

Millon Behavioral Health Inventory (MBHI). The Millon Behavioral Health Inventory (MBHI; Millon, Green, & Meagher, 1979) was designed to measure people's response to medical evaluation and treatment. The MBHI consists of eight scales which assess coping styles in the medical setting as well as 14 other scales measuring psychogenic attitudes, somatization, and prognoses. The eight coping styles on which this study focuses include Introversive, Inhibited, Cooperative, Sociable, Confident, Forceful, Respectful, and Sensitive (see Table 1 for descriptions). The MBHI appears to be a valid and reliable instrument (Millon, Green, & Meagher, 1982). The factor scores produced by the MBHI factor analysis will be used to represent coping styles in this study.

Treatment Compliance/Collaboration Rating Scales. These rating scales were developed in order to measure the level of treatment compliance, interpersonal rapport, alliance, and collaboration between the therapist and the patient in a multidisciplinary pain treatment setting. No other instrument of this kind has yet been developed. Domains of the treatment compliance/collaboration rating are pain management, relaxation, emotional management, activity management, social functional restoration, recreational functional restoration, vocational functional restoration, substance/medication management, weight management, and autonomic nervous system management/neuromuscular re-education (see Appendix A). Domains of compliance/collaboration are rated by the patient's attending psychologist on a 5-point scale ranging from Needs Improvement to Self-Directed. An Overall Compliance Score is computed by adding the 10 ratings and dividing by the number of domains rated (e.g. excluding "not applicable"). For a sample of 31 patients, the median inter-rater reliability for the overall compliance score was found to be .87 among three raters (therapists).

Multidimensional Pain Inventory (MPI). The West Haven-Yale Multidimensional Pain Inventory (MPI; Kerns, Turk, & Rudy, 1985), as described in Study I, is a comprehensive, psychometrically sound instrument which is composed of three sections with a total of 13 empirically derived scales (Kerns et al., 1985). The present study focuses on only one of the scales, Interference. The Interference scale assesses the patient's perception of how much and in what ways the patient perceives his/her pain to affect daily functioning, and thus will be used to represent functional impairment. The means and standard deviation for the Interference scale among a sample of chronic pain patients are M=55.66, SD=7.98. Higher numbers are indicative of more functional impairment. The MPI is a reliable and valid instrument (Jamison, Rudy, Penzien, & Mosley, 1994). An improvement score was generated for each patient by subtracting the pre-treatment Interference score from the post-treatment Interference score. Thus, negative scores are indicative of improvement from pre-treatment to post-treatment.

Procedure

The patients receiving treatment at the pain center completed the MBHI during their first visit. After completing 18-22 sessions of cognitive-behavioral therapy within a 6-month period of time, the patients were administered the MPI. Within two months after patients completed treatment, their attending psychologist completed a Treatment Compliance Rating Scale.

Data Analysis

The factors retained from the factor analysis represented the classifications of coping styles in the current path model. First, these factors were correlated with treatment compliance ratings and improvement in functional capacity. A path analysis was then conducted to obtain direct and indirect effects between the variables, allowance for error terms (e.g. measurement error), a model R^2, and an indication of overall "fit" of this model. Path analysis also allowed for multiple dependent variables in one path model (as compared to multiple regression, which only allows one dependent variable at a time to be analyzed). Path analysis allowed for a graphical representation of relationships between variables, as represented by path coefficients. Path coefficients are either Pearson correlation coefficients or beta weights, depending upon the number of variables predicting the endogenous (dependent) variable (Schumacker & Lomax, 1996). Model fit indices yielded a) the difference between the path coefficients and original (correlation) coefficients among the variables (thus indicating any under/overidentification of the model); and b) the likelihood that this model will replicate across different samples of chronic pain patients. The path diagram, as illustrated in Figure 1, shows treatment compliance/collaboration as hypothesized to be the mediating variable between coping styles and treatment outcome. The R^2 for predicting treatment compliance was .375 and the R^2 for predicting functional impairment was .274. The R^2 for the path model is therefore equal to:

$1 - (1 - .375)(1 - .274) = .546$.

Results

Correlational analyses revealed Factor One (Expression of Negative Emotion) to be positively related to compliance, whereas Factor Two (Aggression) was negatively related to compliance (Table 3). Compliance was positively related to reductions in functional impair-

Table 2. Means for Treatment Outcome Variables

	Mean	Standard Deviation
Factor1	0.03	1.02
Factor2	0.02	1.04
Compliance	3.70	0.62
Interference (Improvement)	-1.44	1.17

Table 3. Correlations Among Treatment Outcome Variables (*N*=67)

	Improvement	Compliance	Factor1
Improvement	1.00		
Compliance	-0.21	1.00	
Factor1	-0.04	0.25	1.00
Factor2	-0.09	-0.33 *	-0.06

Note. * indicate values with probability of p <.05.

ment. As shown in Table 4, the fit indices for this model appear to be a good model fit. However, correlations between compliance and the other variables may be underestimated due to the small variance associated with compliance (see Table 2). The lowest compliance rating given a patient was a three (out of five points). Thus, most patients in this study were rated as having at least satisfactory overall treatment compliance.

Discussion

These results lend support for a mediational model of treatment outcome in a pain management center. Coping styles predict the manner in which patients comply with treatment, and compliance predicts patients' improvements in functional capacity. Results indicate that amplification of emotional distress leads to less compliance with treatment, resulting in poorer outcome. The more emotionally constrained or stabilized, the more compliance patients exhibit and in turn, the higher improvements they attain in functional capacity. Likewise, the more aggressive and forceful patients are in their coping styles, the less likely they are to comply and respond to treatment.

Thus, it appears that possessing a defensive, Polyannish style of coping is much more advantageous in terms of complying with treatment and having a positive treatment outcome. Undergoing emotional distress, depression, and/or other psychopathology, coupled with a lack of defensiveness, appears to put patients at risk for not complying with treatment, and in turn, having a poor treatment outcome. Moreover, being forceful, having Type A personality traits, and being aggressive is also a detriment to treatment compliance and outcome.

The confirmation of this model emphasizes the importance of treatment compliance in MPCs. Moreover, the psychologist-rated compliance scales appeared to be useful as a measure of treatment compliance as well as a predictor of treatment outcome. Compliance appears to be the link between coping/personality styles that patients possess when entering into treatment, and the improvement they've accomplished by the end of treatment. These findings confirm that of other studies using the MBHI as predictors of compliance in health care settings (Tracy et al, 1988).

Conclusions

Findings from Study II indicate that having a coping style that involves suppression and denial of negative emotion facilitates compliance with treatment. However, these findings are not intended to suggest that suppressing negative emotion is *functional*. Possessing defensive coping traits (e.g. being emotionally constrained/stable, Polyannish) can be healthy when one is living a relatively stress-free life. However, when the non-expressive person is faced with a severe stressor that does not go away, such as a chronic pain disorder, denying emotional distress and being defensive may become maladaptive (Wickramasekera, 1993). . This phenomenon has been evidenced in the study of End State Renal Disease patients. Social withdrawal and social alienation were found to be significantly related to poor compliance and poor prognosis (Tracy et al., 1987). Likewise, in a study by Esterling et al. (1990), those chronic pain patients who were repressors, were non-expressive, and disclosed little about themselves were found to have the lowest levels of immune functioning. Defensiveness, which is closely related to avoidance and non-disclosure, has also been found to be related to lower levels of immune functioning (Jamner et al, 1988). Consequently, while being on the non-expressive end may appear to be better than being on the amplifying end, both are likely to be dysfunctional for patients in the long run.

Expression of Negative Emotion and Aggression are, by and large, orthogonal factors. Patients scoring either high or low on Expression of Negative Emotion can score either high or low on Aggression. Judging from the path analytic results, it is most desirable to score on the repressive end of the Expression of Negative Emotion factor and the passive end of the Aggression factor. These patients are likely to be most compliant with treatment and exhibit the most treatment improvements. The most difficult patients are most likely those who score on the amplifying end of Repression/Amplification and the aggressive end of Aggression. Not only are these patients suffering from high levels of emotional distress and functional impairment, but they are also hostile, resentful and aggressive in their approach to treatment. These patients are likely to be most difficult to work with and have a smaller chance of completing a treatment program.

Address correspondence to:
Daisha J. Cipher, Ph.D.
Southern Methodist University
Dallas, Texas 75275-0500

Table 4. Goodness of Fit Criteria for Path Model of Treatment Outcome.

Criterion	Value	Acceptable Level*
Chi-square	3.04	Tabled Chi-square value
GFI (Goodness of fit)	0.98	0 (no fit) to 1 (perfect fit)
AGFI (Adjusted GFI)	0.92	0 (no fit) to 1 (perfect fit)
RMSEA (Root-mean-square error of approximation)	0.01	<.05
AIC (Akaike information criterion)	17.04	Negative value = poor fit

Note. * based on Schumaker & Lomax, 1996.

References

Bonica, J.J. Importance of the problem. In S. Anderson, M. M. Bond, M. Mehta and M. Swerdlow (Eds.), *Chronic Non-cancer Pain*, MTP Press Limited, Lancaster, UK, 1987, p.13.

Cipher, D.J. (1999). Factors affecting treatment outcome and cost effectiveness of multidisciplinary pain management. *Doctoral Dissertation*.

Cipher, D.J., & Clifford, P.A. (1996). Somatization processes in chronic pain patients: MMPI, MBHI, MPI, CIPI, and BDI profiles. *Texas Psychological Association Annual Convention, Dallas, Texas*.

Clifford, P.A., Cipher, D.C., Riggsby, L., & Snider R. (1997). *Clinical ratings of psychological care.* Unpublished instrument.

Cox, D., Tisdelle, D., & Culbert, J.P (1988). Increasing adherence to behavioral homework assign-ments. *Journal of Behavioral Medicine, 11(5)*, 519-522.

Davis, P.J., & Schwartz, G.E. (1987). Repression and the inaccessibility of affective memories. *Journal of Personality and Social Personality, 52(1)*, 155-162.

Dickson, L.R., Hays, L.R., Kaplan, C., Scherl, E., Abbott, S., & Schmitt, F. (1992). Psychological profile of somatizing patients attending the integrative clinic. *International Journal of Psychiatry in Medicine, 22(2)*, 141-153.

Esterling, B.A., Antoni, M.H., Kumar, M., & Schneiderman, N. (1990). Emotional repression, stress disclosure responses, and Epstein-Barr viral capsid antigen titers. *Psychosomatic Medicine, 52*, 397-410.

Flor, H., Fydrick, T. & Turk, D.C. (1992). Efficacy of multidisciplinary pain treatment centers: A meta-analytic review. *Pain, 49*, 221-230.

Gatchel, R.J., Deckel, A.W., Weinberg, N., & Smith, J.E. (1985). The utility of the Millon Behavioral Health Inventory in the study of chronic headaches. *Headache, 25*, 49-54.

Gatchel, R.J., Mayer, T.G., Capra, P., Diamond, P., & Barnett, T. (1986). Quantification of lumbar function. Part 6: The use of psychological measures in guiding physical functional restoration. *Spine, 10*, 36-42.

Gatchel, R.J., Polatin, P.B., Mayer, T.G., & Garcy, P.D. (1994). Psychopathology and the rehabilitation of patients with chronic low back pain disability. *Archives of Physical Medicine Rehabiliation, 75*, 666-670.

Goldstein, D.A., & Antoni, M.H. (1989). The distribution of repressive coping styles among non-metastatic and metastatic breast cancer patients as compared to non-cancer controls. *Psychology and Health, 3*, 245-258.

Jamison, R.N., Rudy, T.E., Penzien, D.B., & Mosley, T.H. (1994). Cognitive-behavioral classifications of chronic pain: Replication and extension of empirically derived patient profiles. *Pain, 57*, 277-292.

Jamner, L.D., Schwartz, G.E., & Leigh, H.(1988). The relationship between repressive and defensive coping styles and monocyte, eosinophile, and serum glucose levels: Support for the opiod peptide hypothesis of repression. *Psychosomatic Medicine, 50*, 567-575.

Kerns, R.D., Turk, D.C., & Rudy, T. E. (1985). The West-Haven Yale Multidimensional Pain Inventory (WHYMPI). *Pain, 23*, 345-356.

Marron, J.T., Fromm, B.S., Snyder, V.L., & Greenberg, D.B. (1984). Use of psychologic testing in charac-terizing the frequent user of ambulatory health care ser-vices. *The Journal of Family Practice, 19(6)*, 802-806.

Millon, T., Green, C.J., & Meagher, R.B. (1979). The MBHI: A new inventory for the psychodiagnostician in medical settings. *Professional Psychology, 10*, 529-539.

Millon, T., Green, C.J., & Meagher, R.B. (1982). A new psychodiagnostic tool for clients in rehabilitation settings: The MBHI. *Rehabilitation Psychology, 27(1)*, 23-35.

Schumacker, R.E., & Lomax, R.G., (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, New Jersey: Lawrence Erlbaum.

Tracy, H.M., Green, C., & McCleary, J. (1987). Noncompliance in hemodialysis patients as measured with the MBHI. *Psychology and Health, 1*, 411-423.

Turk, D.C. (1996). Efficacy of multidisciplinary pain centers in the treatment of chronic pain. In: J.N. Campbell and M.J. Cohen (eds.) *Pain Treatment Centers at the Crossroads: A Practical Conceptual Reappraisal.* Seattle, Washington: IASP Press.

Weinberger, D.A., Schwartz, G.E., & Davidson, R.J. (1979). Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 88*, 369-380.

Weisberg, M.B., & Page, S. (1988). Millon Behavioral Health Inventory and perceived efficacy of home and hospital dialysis. *Journal of Social and Clinical Psychology, 6(3/4)*, 408-422.

Wickramasekera, I. (1993). Assessment and treatment of somatization disorders: The high risk model of threat perception. *Handbook of Clinical Hypnosis* (pp. 587-621). Washington, D.C.: American Psychological Association.

Wilcoxson, M.A., Zook, A., & Zarski, J.J. (1988). Predicting behavioral outcomes with two psychological assessment methods in an outpatient pain management program. *Psychology and Health, 2*, 319-333.

Wothke, W., Arbuckle, J. (1997). *Structual Equation Modeling Workshop Using Amos*. Chicago, Illinois.