Methodological and epistemological issues on linear regression applied to psychometric variables in problem solving: rethinking variance

Dimitrios Stamovlasis, Aristotle University, Greece

ABSTRACT. The aim of the present paper is two-fold. First, it attempts to support previous findings on the role of some psychometric variables, such as, M-capacity, the degree of field dependence-independence, logical thinking and the mobility-fixity dimension, on students’ achievement in chemistry problem solving. Second, the paper aims to raise some methodological and epistemological issues concerning the implementation of the general linear model (GLM) in this type of research. Multiple regression analysis was used to analyze the data, which were taken from students (N =86) in tenth grade of high school taking a compulsory course in chemistry. Three different techniques were implemented in order to support a linear model: the Added Variable Plots, the Stepwise Regression and the Best Subsets Regression. Residual analysis and collinearity diagnosis were also performed in order to test the robustness of inferential statistics. The GLM explained 39% of the variance and suggested that only M-capacity and logical thinking were the significant predictors, even though all the correlation coefficients with achievement were statistically significant. The extensive analysis of the linear regression procedures revealed their advantages and also their limitations in terms of statistical robustness. Moreover, a discussion is initiated concerning the explanatory power of linear models and suggests rethinking variance explained under a different philosophical perspective. It is argued that the weakness of the GLM in studying complex dynamical processes, such as problem solving, is rooted not merely in the statistical assumptions that do not hold, or in the variables that are ignored, but substantially it is deeply epistemological.

KEYWORDS: Cognitive variables, General linear model, Mental capacity, Mobility-fixity dimension, Multiple regression analysis
Educational science, and social science in general, have developed theories grounded on experimental data. Thus, methodologies used in research and the tools implemented for analyzing data have certainly shaped the theoretical development. In educational research the quantitative approaches mainly employ inferential statistics as a tool to draw conclusions and make generalized statements. Contrary to physical science, the application of stochastic methods in social and behavioral sciences faces a great challenge, that is, how to control a plethora of variables and a relatively large number of variability sources, which in many cases remain unknown. The statistical methods attempt to explain the variability by choosing and testing the appropriate variables within a model, while the unexplained portion of variance is treated as noise. Experimental data, that is a sample from a population, are collected and are analyzed statistically in order to determine the extent to which certain patterns observed in the data are generalizable for the population. In social science statistical techniques used in research include simple comparison tests, such as $t$-tests, $\chi^2$, and more complex ones, such as regression analysis (RA), analysis of variance (ANOVA), factor analysis (FA), and structural equation modeling (SEM), to name those best known. However, in the science education literature, and especially in chemistry education research, often findings are reported without using rigorous statistical techniques. This might raise questions about the generalized conclusions drawn for theory development, and suggest further examinations with more robust methodological approaches.

An appealing and motivating area of investigation is the role of individual differences and particularly the role of psychometric or cognitive variables in problem solving or in students’ general performance. Such variables are the information processing capacity (working-memory capacity and/or M-capacity), disembedding ability, logical thinking (developmental level) and the mobility-fixity dimension. Disembedding ability refers to the degree of field dependence/field independence, and represents the ability of a subject to discern information in a variety of complex and potentially misleading instructional contexts (Witkin et al., 1974; Pascual-Leone, 1989). Logical thinking or developmental level is a Piagetian concept, and refers to the ability of the subject to use formal reasoning (Lawson, 1978, 1985, 1993). These have been investigated for more than two

The above cognitive variables are associated with performance in various mental tasks through psychological theories. For instance, the construct of working-memory capacity is part of Baddeley’s model of information processing (Baddeley, 1990). Field dependence-independence has also been associated with information processing as a moderator variable. $M$-capacity is associated with information processing within the neo-Piagetian theory of Pascual-Leone (1970) known as the Theory of Constructive Operators (TCO). According to TCO, the execution of a cognitive task is accomplished by the act of mental operators, such as the $M$-operator associated with information processing, the $L$-operator associated with logical operators and formal reasoning, the $F$-operator associated with field dependence-independence and so on. The aforementioned psychometric variables are measures of the capacity of such mental operators.

The mobility-fixity dimension is in fact a cognitive style associated with the theories of Werner (1957), Witkin and Goodenough (1981) and Pascual-Leone (1989). It is a bipolar dimension, which categorizes subjects as fixed or as mobile. According to Pascual-Leone, field independent subjects are individuals for whom the “overcoming process”, i.e. “strategy x”, is stronger than the “embedding context” created by “strategy y”; the converse is true for field dependent subjects. The scores obtained by subjects on measures of field-dependence-independence (FDI) are a function of the weight of strategy x, relative to the weight of strategy y; as a result of this, the same scores on measures of FDI can be obtained by subjects with vastly different absolute weights in strategies x and y. Pascual-Leone suggests that “Mobile FI subjects are precisely those who exhibit both a very high absolute weight overcoming process and a very high absolute weight embedding context. This expectation implies, of course, that the subjects’ scores on FDI measures will show them to be moderate (but not extreme) FI” (Pascual-Leone, 1989, p. 33).

More about the theoretical aspects of the cognitive variables in science education literature could be found in past papers...

From the methodological point of view, researchers have been using mainly the general linear model (GLM) approach, and in most cases, statistical inference has been based on simple comparison tests and correlation analysis, which is the simplest form of applied regression for two variables.

This research report is part of an investigation on cognitive variables and aims to provide further empirical evidence on their effect. However, primarily it aims to draw attention to methodological issues concerning the implementation of linear regression and to some problems that might lead to a sub-optimal analysis. Moreover, a discussion is initiated concerning the limitations of the GLM in relation to some epistemological issues.

Some methodological issues on linear regression analysis

When more than one independent variable can be potential predictors, one may incorporate all of them in a stochastic equation, such as equation (1), and test its significance by means of linear regression (Ordinary Least Squares, OLS; Draper, Smith, 1981).

\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i + \varepsilon_i
\]  

(1)

There are some issues to be concerned about when selecting this method: the dependent variable has to be measured at least at the interval level, and be unbounded, that is, the observed values exist in the entire spectrum of the scale and they are not constrained within a small range. The predictors are naturally associated with the response, that is, their choice is theory driven. In addition the “third-variable” problem should be addressed. This concerns the existence of a third measured or unmeasured variable that might affect the results (Field, 2005). It is important to stress also that the issue of causality is not addressed statistically by the stochastic equation (1). The betas in the regression model, which represent the rate of change in the response per 1-unit increase of the independent variable, do not imply causal links. The “effect” and the direction of causality are implicitly stated by the researcher.
Moreover, there are two additional aspects to be mentioned here, which will also be discussed later. The first concerns the assumption on which the linear regression analysis is based. Given that all predictors have indeed a linear relation with the response (which is actually being tested), there are some assumptions that must be met so that the estimated parameters of the model, the coefficient betas, \((b)\), are BLUE (Best Linear Unbiased Estimators). The assumptions are that error term \(\varepsilon_j\) has to follow normal distribution \([\varepsilon_j \sim N (0, \sigma^2)]\) and the independent variables have to be uncorrelated with the error term; \(\varepsilon, \beta\) have to be identically and independently normally distributed errors (i.i.d).

The second aspect concerns the utilization of the least-squares method and the choice of the appropriate procedure. In practice, the predictors and the way in which they are entered into the model could have a great impact on the analysis (Field, 2005). There are various approaches, such as hierarchical entry, forced entry, stepwise procedures, and some combined methods that allow an optimal analysis. All many-step procedures have disadvantages due to their ineffectiveness in checking the intermediate residuals and the type I errors. Thus, before one concludes and proposes a model, it is recommended to employ as many procedures as possible. In this work, three different techniques, the Added Variable Plot, Stepwise Regression and Best Subsets Regression were employed. Stepwise Regression is a known algorithmic procedure where variables are kept in the model if it satisfies a predetermined condition (e.g. \(F\)-value > 4 or \(p<0.05\)).

When there are \(k\) independent variables for the model, then \(2^k\) linear sub-models exist. When all models are run then one can proceed by selecting the “best” model based on criteria, such as \(R^2\), \(R^2_{adj}\), \(F\)-values, \(p\)-values, Standard Errors and \(Cp\). The Mallows-Cp is a selection criterion, where the best model will have a value \(Cp \approx k\), where \(k\) is the number of the independent variables in the model\(^1\). The above procedure is the Best Subsets Regression method. An interesting and simple way to judge whether a candidate variable could be introduced in a model is the Added Variable Plot (AVP). Suppose that one examines the model \(Y=f(M, L)\), that is the effect of two independent variables \(M\) and \(L\) on the dependent variable \(Y\). Having run the regression of \(Y\) on \(M\), and consider the model \(Y=f(M)\), the question is, whether \(L\) could enter the model (explaining \(Y\)) given that \(M\) already operates within the model. The regression

1. Mallows– Cp:
\[ Cp : C_p = \left( \frac{SSE}{\sigma^2} \right) + 2p - n \]
where \(SSE\) is the sum of square errors, \(\sigma^2\) is the estimated variance for the complete model \((k\) variables) and \(p\) the number of variables in the model under examination.
2. The regression of $Y$ on $M$ is given by the equation
$$\hat{Y} = \hat{\beta}_{YM} M.$$ 
The residuals of this regression are
$$e_{YM} = Y - \hat{Y} = Y - \hat{\beta}_{YM} M.$$ 

3. The regression of $L$ on $M$ is given by the equation
$$L = \hat{\beta}_{LM} M.$$ 
The residuals of this regression are
$$e_{LM} = L - \hat{L} = L - \hat{\beta}_{LM} M,$$ 
which are values of $L$ with effect of $M$ wiped out.

The symbols with 'hat' refer to the estimated values.

of $Y$ on $L$ (only) that examines the marginal relation of the two variables and it is expressed with the correlation coefficient $r_{Y,L}$, ignores entirely the effect of $M$. Note that the Pearson correlation coefficient is the slope of the regression equation when the two variables are in standard units.

Consider the regression of $Y$ on $M$. The residuals, $e_{YM}$ of this regression are in fact values of $Y$, in which the (linear) effect of $M$ has been erased. Similarly, the residuals $e_{LM}$ of the regression of $L$ on $M$ are values of $L$ with the effect of $M$ erased. The plot $e_{YM}$ versus $e_{LM}$ is the Added Variable Plot (AVP), which depicts the linear relation between $Y$ and $L$ adapted for the variable $M$. The slope of the graph gives the partial correlation coefficient $r_{Y,L,M}$, that is, the correlation coefficient between $Y$ and $L$ given that the $M$ is present in the model. If $r_{Y,L,M}$ is statistically significant then $L$ should enter the model with $M$ already being there. The same rationale could be generalized for the case of three-variable model.

**Method**

**Rationale and hypotheses**

This work is part of a research investigation on cognitive and affective variables that aims to provide empirical evidence on their effect on students’ achievement in science and mathematics education. It emphasizes methodological issues by providing an extended analysis of linear regression (GLM) and it aims to draw attention to problems arising when linear statistics are employed. Thus, the research questions and hypotheses concern the significance of the effect of the four psychometric variables, $M$-capacity, the degree of field dependence-independence, logical thinking and the mobility-fixity dimension, given that all variables are present in the given model.

**Data and measurements**

The subjects ($N = 86$, 48% male) were students in the 10th grade of high school, age 16, who were taking (among others) two compulsory courses, one in physics and one in chemistry. The data for chemistry problem solving are reported here. The sample consisted of students from two different schools located in the broad area of Athens. Data were collected during one school year through paper-pencil test. Students’ achievement scores used as
the dependent variable were from a problem-solving task, which was part of their final examination.

The problem: in 400 mL aqueous solution of NaOH 20% w/v we added 100 mL of water. Calculate the number of moles of NaOH in the final solution [The atomic masses were given as: H=1, O=16, Na=23]. The particular problem was selected to assess students’ understanding the solution chemistry taught during the semester then current. According to what has been taught this problem is not very demanding, and it is rather an algorithmic one, given that the students were supposed to have been practising similar problems. The redundant data of the 100 mL water addition was included as misleading information to reveal “raw” or “parrot” learning. The range of students’ scores was 2.3-10 (min 0, max 10) and the mean score was 7.62 (SD = 2.56). All students were assessed for the following four psychometric variables:

M-capacity: was assessed by means of the Figural Intersection Test (Pascual-Leone, Burtis, 1974). The test was used for measuring the functional M-capacity of the subjects according to the method of Niaz (Pascual-Leone, Burtis, 1974; Niaz 1988a). Cronbach’s alpha reliability coefficient for the present sample was 0.88. The range of students’ scores was 2.5-7.0 (min 0, max 9) and a mean functional M-capacity of 5.10 (SD = 1.22) was calculated.

Cognitive style / Disembedding ability: disembedding ability is usually assessed by means of the Group Embedded Figures Test, GEFT; however, a similar test, the Hidden-Figures Test (HFT) was used, which was devised and calibrated by Johnstone’s group (Johnstone, El-Banna, 1987) from Witkin’s original test materials. This test had a scale from zero to 18. The range of students’ scores was 2.0-15.0 (min 0, max 18) and the mean was calculated 9.87 (SD=2.87). Cronbach’s alpha reliability coefficient was 0.86.

Logical thinking (or developmental level): it was assessed by the Lawson test of formal reasoning (Lawson, 1978). The range of students’ scores was 10-100 (min 0, max 100) and the mean was calculated 59.3 (SD=28.9). A Cronbach’s alpha reliability coefficient of 0.83 was obtained.
Mobility-fixity dimension: For the mobility-fixity dimension, the procedure for classifying subjects as mobile or fixed was the same as that used in previous studies. Field-independent students who obtain high score on FIT could be classified as Fixed, as they consistently demonstrate characteristics of field-independence. Similarly those field-dependent students who obtain low scores on FIT could also be classified as Fixed.

Details for this procedure, which involves both the Figural Intersection Test (FIT) and the Group Embedded Figures Test (GEFT), could be found in earlier studies (Niaz, 1989b; Niaz, Saud de Nunez, 1991; Niaz et al., 2000; Stamovlasis et al., 2002).

\[
\begin{array}{cccc}
 & 1 & 2 & 3 & 4 \\
 GEFT (F) & 1 & & & \\
 Fit (M) & 0.43* & 1 & & \\
 Lawson (L) & 0.41* & 0.54* & 1 & \\
 Achievement (Y) & 0.26** & 0.51* & 0.59* & 1 \\
\end{array}
\]

\*p < 0.0001, \**p < 0.05.

Table 1. Correlation matrix of the three cognitive variables and achievement

<table>
<thead>
<tr>
<th>Entry Variable</th>
<th>SS</th>
<th>Entry</th>
<th>SS</th>
<th>Entry</th>
<th>SS</th>
<th>Entry</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>38.91 M</td>
<td>144.58 L</td>
<td>196.69 M</td>
<td>144.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>158.03 F</td>
<td>1.33 F</td>
<td>0.25 L</td>
<td>79.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>28.09 L</td>
<td>79.12 M</td>
<td>28.09 F</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SSRegr 225.03 SSRegr 225.03 SSRegr 225.03 SSRegr 225.03

Table 2. Sequential ANOVA

Those field-independent students who obtained low scores on the FIT were classified as Mobile, as they show diversity in their modes of functioning. Similarly those field-dependent students who obtained high scores on the FIT could also be classified as Mobile.

\[\text{Statistical analysis and results}\]

The correlation matrix represented in Table 1 shows that all psychometric variables have positive and statistically significant linear correlations with achievement. However, Pearson correlation on its own might not be adequate method to judge the effect of an independent variable when there are other variables affecting the response simultaneously. Note, that all psychometric variables are positively correlated with each other as well. This may imply collinearity, which is addressed below.
4. Variance Inflation Factor (VIF) and the tolerance are measures of multicollinearity. Tolerance equals $1/VIF$. A VIF value greater than 1 or a tolerance value below 1 indicates that multicollinearity may be biasing the regression model.

Table 3. Collinearity diagnosis (Dependent variable: Achievement)

<table>
<thead>
<tr>
<th>Model</th>
<th>Dimension</th>
<th>Eigenvalue</th>
<th>Condition index</th>
<th>Variance proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3.845</td>
<td>1.000</td>
<td>0.00 0.01 0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.088</td>
<td>6.602</td>
<td>0.10</td>
<td>0.1 0.82 0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.042</td>
<td>9.528</td>
<td>0.20</td>
<td>0.11 0.00 0.94</td>
</tr>
<tr>
<td>4</td>
<td>0.024</td>
<td>12.650</td>
<td>0.69</td>
<td>0.88 0.17 0.00</td>
</tr>
</tbody>
</table>

Collinearity diagnosis: using multiple regression analysis, a sequential ANOVA was applied. The method is based on the partition of the sum of squares of regression (SSRegr) into three parts corresponding to the three variables (Draper, Smith, 1981). The hierarchical entry of each variable in the OLS procedure affects its contribution to the total SSRegr, which remains constant. If the three variables are orthogonal, then the partition of the total SSRegr is unique. Table 2 shows that the partition of the total SSRegr varies when the hierarchical entry changes. Thus, multicollinearity is present. However, quantitative measurements showed that it is not large enough to worry about.

Two multicollinearity indexes, the variance inflation factor (VIF) and the tolerance are 1.42 and 0.704 respectively. The above raise a small concern about bias in the regression model. Moreover, Collinearity Diagnosis employing principal component analysis was performed (Field, 2005).

Eigenvalues of the scale, uncentred cross-product matrix, condition indexes and variance proportions are presented in Table 3. Each variable $M$, $F$, and $L$ ($M$-capacity, degree of field dependence...
independence, and logical thinking respectively) should have the largest portion of its variance loaded onto different dimension. For the present case, \( L \) has 82% of its variance onto dimension 2, \( F \) has 94% of its variance onto dimension 3 and \( M \) has 88% of its variance onto dimension 4. Table 3 suggests that no serious bias-problems are expected from multicollinearity.

Multiple regression analysis: the objective is to test whether each independent variable affects significantly the achievement given that all variables are present. In other words the following model is to be tested by means of Ordinary Least Squares (OLS).

\[
Y = \beta_0 + \beta_M M + \beta_F F + \beta_L L + \varepsilon_j
\]  

(2)

The issue of statistical power for OLS is worth discussing here. Statistical power depends on the intended effect size, \( R^2 \), the number of variables in the model and the degree of collinearity. Studies on OLS methods suggested that 55 observations would be the minimum appropriate sample size for a three-parameter model and for medium effect size with a power of 0.80 (Maxwell, 2000). Thus, the sample size of 86 subjects is satisfactory, given that no missing values existed in the data.

In equation (2), let \( M \) be the first variable placed in the model. This is a theory driven decision because the information processing capacity plays often the main role in problem solving.

Figure 2. Added Variable Plot: the residuals \( e_{YM} \) versus the residuals \( e_{LM} \). The slope of the regression equation is statistically different from zero. \( b=0.422\pm0.094, t=4.48, p <0.0001, F\text{-value}=20.1, p <0.0001\). Then \( L \) should be introduced in the model.
The next step is to examine the Added Variable Plot in order to test whether $F$ could be introduced in the model. The rationale was explained in the previous section. Fig. 1 is the corresponding AV Plot, that is the plot of the residuals $e_{YM} = Y - \hat{\beta}_{YM}M$ versus the residuals $e_{FM} = F - \hat{\beta}_{FM}M$. The slope of the AVP is not statistically different from zero. \[b=0.048 \pm 0.093, \ t=0.52, \ p=0.603, \ CI(95\%)=\left[-0.136, 0.233\right], \ F\text{-value}=0.27, \ p=0.603\]. This suggests that $F$ should not be introduced in the model.

The next step is to test whether $L$ could be introduced in the model given that $M$ is present. Fig. 2 is the corresponding AV Plot of the residuals $e_{YM} = Y - \hat{\beta}_{YM}M$ versus the residuals $e_{LM} = Y - \hat{\beta}_{LM}M$. The slope of this AV Plot is statistically different from zero. \[b=0.422 \pm 0.094, \ t=4.48, \ p<0.0001, \ F\text{-value}=20.1, \ p<0.0001\]. This suggests that $L$ should be introduced in the model.

Finally, the AV Plot of Fig. 3, that is, the graph of the residuals $e_{YL,M} = Y - (\hat{\beta}_{YL} + \hat{\beta}_{YMM})$ versus the residuals $e_{FL,M} = F - (\hat{\beta}_{FL} + \hat{\beta}_{FYLM})$ will provide the information needed to decide whether $F$ should be placed in the model, given that both $M$ and $L$ are present. The slope of the regression equation in Fig. 3 is not statistically different from zero \[b=-0.41 \pm 0.086, \ t=-0.482, \ p=0.631, \ CI(95\%)=\left[-0.212, 0.129\right], \ F\text{-value}=0.232, \ p=0.631\]. This suggests again that $F$ should not be introduced in the model.
Stepwise regression supported the above conclusion by rejecting $F$ from the model. Table 4 shows that the regression procedure finally converged with two dependent variables in the model: $M$-capacity and logical thinking.

The best subsets Regression also indicated the same model as the best for the data (Table 5). Examining the criteria, this model has the larger $R^{2}_{adj}$ (38.8%) and the smaller standard error ($s=2.00$) of the regression, while the Mallows- $C_p$ $(=2.2 \approx 2 = \text{the number of independent variables})$ is the optimum.

### Table 4. Stepwise Regression: on the three predictors

<table>
<thead>
<tr>
<th>predictors</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4.047</td>
<td>2.105</td>
<td>2.105</td>
</tr>
<tr>
<td>LAWSON (L)</td>
<td>$\beta$</td>
<td>0.557</td>
<td>0.422</td>
</tr>
<tr>
<td>t-value</td>
<td>6.77</td>
<td>4.45</td>
<td>4.45</td>
</tr>
<tr>
<td>FIT (M)</td>
<td>$\beta$</td>
<td>-</td>
<td>0.55</td>
</tr>
<tr>
<td>t-value</td>
<td>-</td>
<td>2.66</td>
<td>2.66</td>
</tr>
<tr>
<td>GEFT (F)</td>
<td>$\beta$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>t-value</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>s</td>
<td>2.07</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>$R^2$</td>
<td>35.29</td>
<td>40.21</td>
<td>40.21</td>
</tr>
</tbody>
</table>

$F$-to-Enter: 3.84; $F$-to-Remove: 2.71, $N=86$.

### Table 5. Best Subsets Regression

<table>
<thead>
<tr>
<th>Number of Variables</th>
<th>$R^2$</th>
<th>$R^2$-adj.</th>
<th>C-p</th>
<th>s</th>
<th>GEFT (F)</th>
<th>FIT (M)</th>
<th>LAWSON (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.3</td>
<td>34.5</td>
<td>7.0</td>
<td>2.0721</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25.9</td>
<td>25.1</td>
<td>19.9</td>
<td>2.2167</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>40.2</td>
<td>38.8</td>
<td>2.2</td>
<td>2.0037</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>35.3</td>
<td>33.8</td>
<td>8.9</td>
<td>2.0839</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40.4</td>
<td>38.2</td>
<td>4.0</td>
<td>2.0131</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

### Table 6. Regression slopes, Standard Errors, t-tests, Confidence intervals and Model fit for the Linear Model

<table>
<thead>
<tr>
<th>Model</th>
<th>$b$</th>
<th>SE(b)</th>
<th>Standardized $b$</th>
<th>$t$</th>
<th>95% CI</th>
<th>$R^2$</th>
<th>Model F</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAWSON (L)</td>
<td>0.421</td>
<td>0.0947</td>
<td>0.450</td>
<td>4.45**</td>
<td>0.233</td>
<td>0.610</td>
<td>27.9***</td>
</tr>
<tr>
<td>FIT (M)</td>
<td>0.554</td>
<td>0.2117</td>
<td>0.264</td>
<td>2.61*</td>
<td>0.132</td>
<td>0.975</td>
<td></td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; Durbin-Watson 1.75; VIF= 1.42; tolerance=0.704.
To test the effect of the Mobility Fixity dimension ($D$), it was introduced as a dummy variable, where $D=1$ for mobile students and $D=0$ for fixed students.

$$Y = \beta_0 + \beta_M M + \beta_L L + \beta_D D + \varepsilon_j$$  \hspace{1cm} (3)

Again, all regression procedures rejected $D$, while the original model was not improved. Note that for the Mobility-Fixity Dimension nonparametric tests (and the parametric $t$-test) showed that the achievement of mobile students was statistically higher than the corresponding achievement of the fixed students (Mann-Whitney test: $z = -2.98$, $p < 0.01$; Kolmogorov-Smirnov test: $z = 1.697$, $p < 0.01$; $t$-test: $t = -2.77$, $p < 0.01$).

Thus, the final proposed model includes two variables, $M$ and $L$. The model parameters and statistics are shown in Table 6. Note that the values of the standardized betas suggest that the effect of Logical thinking (0.450) is greater than that of $M$-capacity (0.264).

Residual analysis: the basic assumption in linear regression analysis is that error term $\varepsilon_j$ has to follow normal distribution $[\varepsilon_j \sim N (0, \sigma^2)]$. The error term is the unexplained portion of the variance and it is measured by the residuals of the regression equation.

![Histogram of the Residuals](image)

Fig. 4 shows the distribution of the residuals and Fig. 5 shows the normal $P$-$P$ plot, which indicates a deviation from normality. In addition, Kolmogorov-Smirnov and Shapiro-Wilk tests have both rejected the null hypothesis for normality ($p<0.001$). That is, the first assumption is violated. However, the homoskedasticity is satisfied. Fig. 6 shows the residuals versus fitted value. The White
The White test for heteroskedasticity is available in the E-Views software. The Durbin-Watson test statistic, which tests the assumption of independent errors, gives a value of 1.75, which indicates small positive serial correlations between errors, but they are not statistically significant. The assumption of independence was not rejected either by the Run-test ($z=-0.429$, $p=0.688$), which tests serial correlations in a symbolic sequence with a non-parametric procedure.

**Discussion: rethinking variance**

The initial hypothesis that the degree of field dependence-independence and the mobility-fixity dimension affect students’ achievement in the mental task employed in this research is not supported by the data. However, the hypothesis was supported for logical thinking and $M$-capacity. The former, representing formal reasoning, is required in learning science and the latter is associated with information processing. The statistical significance of both variables being present in the model is theoretically expected.
The role of these cognitive variables has also been supported in research concerning molecular equilibrium problems (Tsaparlis et al., 1998). On the other hand, the role of field dependence/independence appears diminished when examined along with the effect of logical thinking and information processing capacity in this algorithmic-type problem.

At this point a methodological assumption should be discussed. When algorithmic problem solving or any assessment test of specific domain in science education is implemented to evaluate learning outcomes, the effect of disembedding ability or other psychometric variable should be mainly considered as a determining factor of the preceding learning process, and not of the assessment process. Otherwise, the instrument would measure more than what is meant to measure and validity issues will be raised. Thus, possible utilization of the redundant data by students in the present study would rather indicate “raw” or “parrot” learning, which in turn, might be considered as the product of the learning/teaching process, where lack of disembedding ability, insufficient formal reasoning, low information processing capacity, or a combined effect had prevented students from reaching conceptual understanding.

When assessing a learning process with cross-sectional data, the protagonist factors which catalyze this process, can only manifest their effect by means of statistics, that is, by explaining portion of the variability of students’ scores. This is the only empirical evidence about the role of these factors that the linear regression model can provide. In this model, not rejecting the null hypothesis that the linear effect of a variable is zero does not exclude any nonlinear/operational role (e.g. threshold effects), given its natural association with the mental process. In algorithmic problem solving, some effects, such as the role of field dependence-independence in this study, appeared diminished because it might be “masked”
or covered by the effect of learning. In a non-algorithmic problem or “real” problem, however, where high order cognitive skills are required, the role of constructive operators, to mention the theoretical constructs within the neo-Piagetian framework, is expected to be more decisive during the assessment process and the statistical effect of some cognitive variables to appear more pronounced. The above could explain research findings reporting that field dependence-independence has an enhanced role in non-algorithmic problem solving, while it is non-significant predictor for algorithmic case (Tsaparlis, 2005).

Researchers state theory-driven hypotheses and they seek to provide empirical evidence for statistically significant effects, which are important to make generalized statements. The statistical method is crucial, however. Correlation analysis might be falsified by multiple regression analysis. The findings confirm that two variables are important, logical thinking and M-capacity and the mental task requires information processing and formal reasoning. The linear model quantifies their effect by explaining 39% of the variance. The residual analysis, however, suggests that the estimated parameters may not be BLUE (Best Linear Unbiased Estimator), because the basic assumption of normality is not satisfied. This affects the inferential power of the statistical method. In educational research such an analysis is usually not reported. In some other social sciences, such as economics, this is a very important issue since the models are used to predict future values, e.g. the next year’s inflation. In educational science such models are not used in the same way. The prediction of an individual’s achievement on the basis of his psychometric scores is not only impossible, but it is not our objective. The aim of modeling the distribution of the response, in this field, is to explain the variability in terms of “predictors”, which are not “predicting” the outcome in the way that it is assumed in other research fields. What a researcher merely attempts to accomplish is to provide empirical evidence through inferential statistics that these variables have a significant contribution. However, in theory building, the statistical methodology has to be robust and this is something that deserves attention.

The GLM has three weak points and limitations. First, it explains only a small portion of the variability. Second, it has limited inferential power resulting from the violation of the basic assumptions. Third, it does not provide a meaningful explanation for or description of the
process being studied (problem solving). That is, a linear addition model of cross-sectional data provides only the quantitative average change in the response per 1-unit increase of the independent variable, while no information is given about causality or any clue about how the process is evolving in time.

A question arises about the origin of these limitations. Concerning the small variance explained, one might argue that it is due to measurement error and/or some additional variables, which are ignored. These are both true to some extent. Measurement errors are certainly involved, as they are in any other scientific field. In physics, for instance, when the Ohm’s law was invented in 1826, the errors of measurement were dreadfully large; however the “law”, that is, the linear relationship between the intensity of electric current and the potential difference, was uncontested. Even large measurement errors, as it was shown later with the development of statistics, have to follow Gaussian distribution and Ohm’s law was not challenged. Large measurement errors did not prevent the discovery of a fundamental law in classical physics. Note that this “law” concerns a simple material system.

In social and behavioral sciences, however, things are different and rather more difficult. Checking the error term of a linear stochastic equation disappoints the researcher, because it is large and does not follow the bell-shaped distribution, and he/she often wonders what is missing or what went wrong. The usual act is to seek additional variables for the model. Even if theoretically the more variables in the model the larger the variance explained, does it actually happen in educational research? Suppose that all predictors in question become known, will it be possible to express the outcome of students’ performance as a simple weighted linear sum of the contributing variables? The answer depends on the ontological assumption about the nature of the problem solving process and brain functioning.

The main thesis of the present discussion is that in this type of research, the weak point of GLM as a research tool is rooted in its incompatibility with the nature of the process that it is supposed to describe. The GLM is compatible with the ontological considerations that foster the mechanical/computer metaphor for the mind, and if this consideration does not hold, the linear approach is epistemologically inadequate to access information or knowledge about this system.
Returning to problem solving, a close examination of the mental processes within the framework of neo-Piagetian theories, which this research is related to, sheds light on the nature of these processes. According to this theory, in a cognitive task execution the mind activates certain operative schemes, which are responsible for the transformation and coordination of the preexisting or the new information (or schemes/knowledge) with the action of variety of constructive operators, each of which performs a specific class of function (Pascual-Leone, 1970, 1974, 1989). At the level of complexity at which science education research studies the cognitive process, one could recognize a number of mental subprocesses. These mental subprocesses correspond to the action of the mental operators employed in neo-Piagetian theories and they are operationalized by the variables implemented in the present research. Thus, in a mental task execution, the mind might proceed as follows: inputs data, retrieves information from long term-memory, processes information, separates “signal” from “noise” (if any), applies formal reasoning, processes information again and so on. The sequence of the above steps, especially in a ‘real’ problem situation, is not predetermined. The mind might activate successively the above subprocesses in a pattern that is not linear or cyclical, but it might be seemingly random, which finally converges to a solution, if any. The mental process does not follow a unique path, but each step is determined by the previous steps. This makes problem solving a dynamical process.

One more issue worth discussing at this point is related to the demand and the nature of the problem. In the present research the demand of the problem is not very high and the problem is not a ‘real’ one. That is, the constructive operators run rather smoothly without having to confront any great challenge. Even so, the GLM, which is compatible with small change situations failed to model adequately the outcome of the process. A question arises as to what happens in the case of a “real” or HOCS (Higher Order Cognitive Skills) type of problem (Tsaparlis, Zoller, 2003), where the constructive operators are probably pushed to their limits. A first thought is that things are getting even harder for the GLM, because stronger deviations from normality in score distributions are expected.

Problem solving is a more complex process when considering the variety of factors involved, such as ability levels, patterns of thinking,
learning styles and personality traits. These components are interacting with each other in task executions in a nonlinear fashion (Stamovlasis, 2006). For instance, field dependence-independence, which is a moderator to information processing, could affect the response below or above a threshold. This effect could not be captured by a linear model. Thus, the rejection of disembedding ability from the linear model does not exclude a possible crucial role in a nonlinear fashion.

This area of investigation concerns mental processes, and could be thought of as the interface between educational science and psychology. It is an applied field that borrows theories from psychological sciences, which in turn, nowadays, owe their development to neuroscience. The latter seeks to understand how brains give rise to thought and behavior and have a long history of mutual influence with psychology (Freeman, 2000). Therefore, advances in psychology and neuroscience have to be taken into consideration when stating hypotheses in educational research, which should be fertilized by new insights into fundamental theoretical issues and be aware about alternative methodological approaches. Some of these issues concern the assumption of linearity and the mechanistic/computer metaphor applied to study brain functioning. The above two issues are interrelated through ontological and epistemological assumptions. The GLM might be a proper tool if the above presuppositions of linearity and mechanistic nature of mental processes are made. However, consolidated knowledge from contemporary neuroscience, which has studied brain functioning as a complex dynamical process, has provided a new paradigm (Freeman, Barrie, 2001). Linear relations in such system are not capable of explaining its function. Ergo, the ontological and epistemological assumptions associated with these phenomena at the behavioral level might be shifted towards nonlinear dynamical systems considerations (Stamovlasis, 2005, 2008).

Having accepted the philosophical assumptions above, one may reconsider the nature of variability and the variance explained. In this study, the GLM left 61% unexplained variance treating it as error. In classical psychometric theory a mental measurement Y consists of a True score, T, and error e. Errors are assumed to be normally distributed and unrelated to each other and to true scores (iid). However, when abandoning the linear approach, the
The inclusion of a nonlinear function is crucial in reformulating the standard measurement theory. In that case, dependent errors are expected to appear in the residuals. It has been shown, that such non-iid errors (residuals) are indicative of nonlinear processes (Brock et al., 1990). The residual analysis in this research suggested that this might be the case. Thus, in nonlinear dynamical processes the score variance is

$$\sigma^2(Y) = \sigma^2(L) + \sigma^2(NL - L) + \sigma^2(de) + \sigma^2(iide)$$

(4)

The four components are the linear, the nonlinear, the dependent errors and the $iid$. A linear model treats the last three components as errors $\sigma^2(e)$, while the dependent errors are captured only by the proper and well-defined nonlinear model and could increase the variance explained (Guastello, 2002).

In conclusion, it is rather difficult for the cognitive variables involved in problem solving to explain the outcome in terms of linear addition. Problem solving is a dynamical mental process where the role and the importance of the involved variables cannot be captured from cross-sectional data. Moreover, a student’s performance that is the outcome of a dynamical process cannot be modeled as the linear sum of all his/her abilities. Variables, such as information processing capacity, logical thinking and field dependence/independence affect the process by acting in each step and interacting with each other and with the problem constraints. Linear treatment of such variables hardly provides a satisfactory description of such mental processes, because it cannot capture their dynamics and thus GLM will always have poor explanatory power. The weakness of the GLM in this domain is rooted not merely to the assumptions that do not hold, or to the variables that are ignored, but substantially it is deeply epistemological.

Rethinking unexplained variance: it might no longer be treated as $iid$ noise. It might include nonlinear components that contain information about the process that the data originate from. Nonlinear models that include, implicit or explicit, the time variable then might be fostered, which would provide a better understanding about problem solving processes and will definitely improve theory building.
Acknowledgement

The author is thankful to the science teachers Mr. C. Modiatis and Mr. M. Brokos, for permitting him to contact the research in their classes, and to Dr. A. Dimos for his valuable help in assessing the psychometric tests.

References


Draper Norman, Smith Harry (1981), Applied regression, New York, John Wiley and Sons

El-Banna H. (1987), The development of a predictive theory of science education based upon information processing hypothesis, Glasgow, UK, University of Glasgow [Doctoral thesis]


Niaz Mansoor, Saud De Nunez G., Ruiz De Pineda Isangela (2000), Academic performance of high school students as a function of mental capacity, cognitive style, mobility-fixity dimension, and creativity, “The Journal of Creative Behavior”, V. 34, pp. 18-29

Pascual-Leone Juan (1989), An organismic process model of Witkin’s field dependence-independence, in Globerson Tamar, Zeiniker Tamar (eds.), Cognitive style and cognitive development, Norwood, NJ, Ablex, pp. 36-70


Pascual-Leone Juan, Burtis P. J. (1974), FIT: Figural Intersection Test, a group measure of M-space, Ontario, York University [Unpublished manuscript]

Scandarmalia Marlene (1977), Information processing capacity and the problem of horizontal decalage: a demonstration using combinatorial reasoning task, “Child Psychology”, V. 48, pp. 301-345


Tsaparlis Georgios, Zoller Uri (2003), *Evaluation of higher vs. lower-order cognitive skills-type examinations in chemistry: implications for university in-class assessment and examinations*, "University Chemistry Education", V. 7, pp. 50-57
Wechsler David (1955), Wechsler adult intelligence scale manual, New York, Psychological Corporation

Werner Heinz (1957), The concept of development from a comparative and organismic point of view, in Harris Dale B. (ed.), The concept of development: an issue in the study of human behavior, Minneapolis, MN, University of Minnesota Press, pp. 125-148


Sintesi

Gli evidenti e centenari successi ottenuti dal metodo scientifico nella ricerca di leggi e nella formulazione di teorie hanno spinto gli studiosi delle scienze sociali, così come psicologi e medici, tanto per fare alcuni esempi, alla raccolta dei dati relativi ai fenomeni di loro competenza al fine di trasportare quel metodo all’interno delle loro discipline.

Nel far questo si è evidenziata una differenza fondamentale tra le materie più propriamente scientifiche e quelle sociali (intendendo con quest’ultime quelle che
hanno l’essere vivente come oggetto): mentre nelle prime le leggi dedotte hanno a che fare con fenomeni generalmente riproducibili, le seconde non godono della medesima fortuna. Questa diversità si può sostanzialmente ricondurre alla presenza dell’essere vivente, oggetto di studio delle scienze sociali ed umane, dell’etologia e della botanica, che può essere descritto soltanto da un numero illimitato di variabili, restando così intrinsecamente sfuggente. In altre parole, per le sociali non vale il determinismo così come è stato efficacemente descritto da Laplace: “Dobbiamo dunque considerare lo stato presente dell’universo come effetto del suo stato anteriore e come causa del suo stato futuro. Un’intelligenza che, per un dato istante, conoscesse tutte le forze da cui la natura è animata e la situazione rispettiva degli esseri che la compongono, se fosse abbastanza vasta da sottoporre questi dati ad analisi abbraccerebbe nella stessa formula i moti dei corpi più grandi dell’universo e quelli dell’atomo più leggero: per essa non ci sarebbe nulla d’incerto, ed il futuro come il passato sarebbe presente ai suoi occhi”.

Questa diversità porta inoltre ad una situazione curiosa: non solo, come detto, l’oggetto misurato è di difficile descrizione e, quindi, di analisi, ma lo è anche il metodo stesso impiegato per tale analisi. Ciò comporta che nello studio dei fenomeni sociali sia necessaria una costante verifica e ricalibrazione degli strumenti, delle metodologie e dei parametri utilizzati.

In questo contesto si inserisce Stamovlasis con il lavoro di analisi del problem solving, in ambito chimico, da parte di studenti tra i quindici e sedici anni. I risultati da lui ottenuti da una parte supportano le scoperte precedentemente raggiunte in questo settore, confermando il ruolo di alcune delle variabili cognitive coinvolte quali la capacità di memoria, il pensiero logico e la dimensione mobilità-fissità; dall’altra però mettono in risalto come i diversi metodi statistici in uso avallino parzialmente e in modo differente questi ruoli. In altre parole, viene evidenziato come ancora non si abbia strumento di analisi sufficientemente generale che, da solo, riesca a giustificare i risultati ottenuti.

In ciò è rilevante il presente contributo: la discrepanza rilevata non dipende tanto dalla metodologia utilizzata quanto dall’approccio epistemologico. I metodi impiegati si basano tutti su modelli lineari, modelli generati a partire dalla metafora meccanicistica che vede la mente come una macchina. Ma se tutti i modelli che ne conseguono non riescono a inquadrare la totalità delle variabili osservate, è probabilmente ciò che ne è all’origine ad essere inadeguato: il problem solving non può essere descritto tramite modelli lineari poiché è un processo dinamico in cui le variabili che ne sono alla base partecipano contemporaneamente sia singolarmente sia influenzandosi vicendevolmente ed è in questa seconda caratteristica che la linearità fallisce.

È il momento di fermarsi. cambiare punto di vista e, coraggiosamente, provare a riscrivere le regole del gioco.