CHAPTER 4:

EMPIRICAL COMPARISON OF NONMETRIC TEMPORAL PATH ANALYSIS AND THE LINEAR MODELS APPROACH

Overview

Education has been defined as a teaching-studenting process in which someone deliberately attempts to guide another who is attempting to learn something somewhere (Steiner, 1978). Thus, the basic components of education are teacher, student, content and setting. The science of education should result in knowledge about these four components and their interrelations. If specific relationships are predictable, this knowledge can be used to forecast results of educational processes and control educational outcomes.

A general systems perspective suggests a state-space approach to the analysis of educational system component relations (e.g., Maccia & Maccia, 1966; Wienberg, 1975). A system has been defined as a group of components with at least one affect relation which has information. An affect relation is a connection of one or more components to one or more other components (Maccia & Maccia, 1966). Based on information theory, states of each system component can be characterized by categories in a classification. The connections among components can be conceived as a subset of the Cartesian product of categories among classifications which temporally characterize changing states of system components. In

NTPA the joint and sequential occurrences of system component states are observed and enumerated. The likelihood of system processes or patterns is estimated by observing an individual system over time and enumerating the relative frequency and/or duration of specified temporal paths. An estimate of the probability of a given relation is obtained for an individual system.

For example, in studying the relation between teacher instruction and student engagement, one might find after observing one classroom on ten different two-hour occasions that the likelihood of student task engagement during direct instruction is .98 for this system. Another independent, individual system might also be observed on ten different occasions, and one might find that the likelihood of the relationship of student task engagement given direct instruction is .94 for this system. Similarly, other independent systems of the same type (e.g., elementary classrooms in the U.S.) may be randomly selected and observed, and the likelihood of this particular relation is estimated for each of these systems.

If it is the case that systems are randomly selected (i.e., each has an equal probability of being selected) from the population of interest and that systems are independent, then a confidence interval that is likely to contain the population mean can be constructed for this particular relationship (student task engagement given direct instruction, or EN|DI). For example, if it can be assumed that the sampling distribution is normal, then a confidence interval can be estimated by:

$$\bar{X} \pm Z_{(\alpha/2)} \text{ S.E.}$$

where $\bar{X}$ is the sample mean, $Z_{(\alpha/2)}$ is the number of standard deviation units above or below the mean of a normal distribution which demarcate a given proportion of its area, and the standard error (S.E.) is the standard error of the sample mean. The standard error decreases as a function of the sample size (N, the number of independent systems observed) and the sample standard deviation (S.D.):

$$S.E. = S.D./(N-1)^{1/2}$$

The alpha level can be chosen for whatever level of confidence desired. These are standard parametric statistical procedures for estimation of confidence intervals (Hays, 1973).

In summary, a relation in NTPA is operationalized as a single variable. The measure of the value of that variable is the probability of occurrence of the nonmetric temporal path which specifies the relation. Replication across independent systems permits averaging of probabilities for individual systems so that statistical generalizations can be made about the mean probability of the relation for the population of systems. In this example, the relation was specified as student engagement given direct instruction (EN|DI) and the measure of the relation was the observed percent time of its occurrence, used to estimate the probability of the occurrence of the relation.

In correlational analysis two variables would be measured for each independent system sampled. A measure of DI and a separate measure of EN would be obtained in this example. To be consistent with the above, these separate measures would be the observed percent time of EN and of DI. The linear relationship of these pairs of measures would be indicated by the Pearson product moment coefficient. The strength of the linear association would be estimated by the square of the product moment coefficient, indicating the percent of variance accounted for in the dependent variable by the independent variable.

In analysis of variance (ANOVA) two treatment groups would be formed: one which received direct instruction and a control group which did not. The dependent variable would be student task engagement. Group means would be statistically compared for differences by means of an F test. Strength of association between independent and dependent variables would be estimated by eta squared or omega squared.

Method

Observational data were collected on 25 mildly mentally handicapped students and their teachers during the second year of a study of academic learning time and student achievement (Rieth & Frick, 1982). Students were observed a total of eight to ten hours each at different times during the school day over a period of about six months. Observational data were collected on coding forms by highly trained observers, using the Academic Learning Time Observation System (ALTOS) (Frick & Rieth, 1981). During mathematics and reading activities observers coded student and instructor behavior at one-minute intervals.

At each sampling moment type of instruction available to the target student was characterized by one of the following categories: Academic monitoring (AM), academic feedback (AF), academic questioning (AQ), academic explaining based on student need (XN), planned academic explaining (XP), academic structuring/directing (SD), academic task engagement feedback (TF), or null (NU). In addition, student academic task engagement was characterized at each sampling moment as: engaged written (EW), engaged oral (EO), engaged covert (EC), engaged with structure/directions (ED), non-engaged interim (NI), non-engaged wait (NW), or non-engaged off-task (NO). These individual categories are defined elsewhere (Frick & Rieth, 1981; Fisher, Berliner, Filby, Marliave, Cohen, Dishaw & Moore, 1978).

For purposes of simplification in the example below, direct instruction (DI) was considered as AM or AF or AQ or XN or XP or SD or TF. Direct instruction refers to academic transaction with the target student or a group of students of which the target student is a member during an educational activity. From the point of view of a given student (i.e., target student), the source of direct instruction could be the teacher, another person in the classroom such as a peer or an aide, or something capable of sending information to <u>and</u> receiving information from the student (e.g., programmed instruction via computer). By and large, the teacher was observed to be the source of direct instruction in the Rieth and Frick (1982) study. If there was no academic transaction with the target student or group containing that student during an academic educational activity, then the type of instruction was considered to be non-direct (NDI). Similarly, student engagement (EN) was considered as EW or EO or EC or ED, and student non-engagement (NE) as NI or NW or NO. A student was considered to be engaged in an academic activity if she or he appeared to be attending to the academic substance of an educational activity. Thus, the two classifications were reduced to two categories each for purpose of illustration:

Classification	Categories
Academic instruction available to student:	Direct (DI) Non-direct (NDI)
Student orientation to instruction:	Engaged (EN) Non-engaged (NE)

Since these two classifications were jointly coded at each sampling moment, a nonmetric temporal path analysis of the relation among categories in these classifications could be performed. Since the data from the Rieth and Frick (1982) study were not collected in continuous real time, true sequential analysis (i.e., 'If ..., then ...' queries--see Chapter 2) was not possible with these data. However, it was possible to analyze the simultaneous occurrence of categories in different classifications.

The Linear Models Approach (LMA): Correlational Analysis

In correlational analysis direct instruction (DI) is considered as a random variable. For each student-instructor pair a measure of DI was constructed by dividing the total frequency of one-minute samples in which DI occurred by the total frequency of one-minute samples observed for that pair. For example, if a student-instructor pair were observed for 600 minutes and 350 were instances of DI, then the proportion of DI, p(DI), is 350/600, or .583. Conversion to proportions was necessary since not all teacher-student pairs were observed for the same total amounts of time. Thus, the proportions normalize the frequencies of one-minute samples to a common metric.

Similarly, student engagement (EN) is considered as a separate random variable, where proportions are likewise used as measures of EN.

Since there are only two categories in each classification, non-direct instruction (NDI) and non-engagement (NE) provide no additional information, respectively, for the instructor and student classifications. For each student-instructor pair the proportion of DI determines the proportion of NDI, and the proportion of EN determines the proportion of NE, since the proportions must sum to one for each classification for each student-instructor pair.

To estimate the relationship of DI and EN a bivariate distribution was formed. See Figure 3. The $\hat{p}$(EN) was .741 (S.D. = .101) and the $\hat{p}$(DI) was .432 (S.D. = .144). The Pearson product moment coefficient was .57 and statistically significant ($p < .05$). Thus, there is a positive relationship between DI and EN, significantly greater than zero, according to correlational analysis, and represented by the linear regression equation, p(EN)' = .57 + .40p(DI). The standard error of beta was .12. Knowledge of the proportion of DI reduces the uncertainty of the prediction of the proportion of EN by about 32 percent ($r^2 = .32$). It would appear that, although positive, the relationship between DI and EN is not very strong in terms of percent of variance accounted for (i.e., proportional reduction of error, or PRE). About twice as much variance is unpredictable as is predictable. The relation is symmetrical, since $r_{EN,DI} = r_{DI,EN}$. Thus, the percent of variance of DI accounted for by EN is also 32 percent. In addition, since the sample size is relatively small, the beta weight for the linear regression equation has a relatively large standard error. One would have little confidence in estimating the population beta with such a small sample.

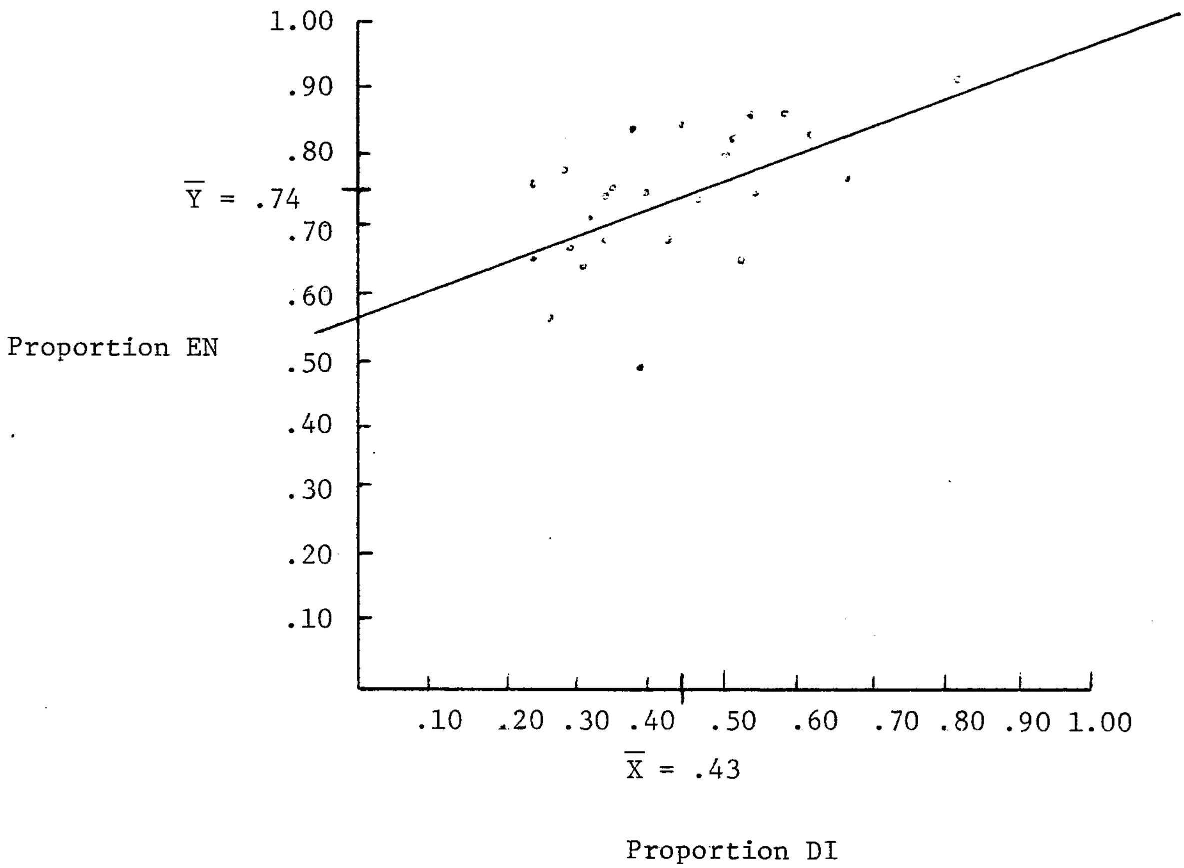


Figure 3. Bivariate Distribution of Direct Instruction (DI) and Student Task Engagement (EN). Regression Line: EN' = .57 + .40DI .

Nonmetric Temporal Path Analysis (NTPA)

The very same observational data for the 25 student-instructor pairs were also analyzed by the NTPA approach. Joint and conditional distributions were formed for DI & EN, DI & NE, NDI & EN, NDI & NE, EN|DI AND EN|NDI. See Table 2. Note that DI & (EN OR NE) = DI; EN & (DI OR NDI) = EN, etc.

The relation, EN|DI, is considered to be a single random variable. For each student-instructor pair a measure of EN|DI was constructed by counting the one-minute samples in which DI & EN occurred and dividing by the frequency of one-minute samples in which DI occurred. The proportion time that this nonmetric temporal path was observed was used to estimate the probability of the relation. The estimated probability of EN given DI, $\hat{p}$(EN|DI), was .967 (S.D. = .029). In other words, given that DI is occurring, students are very likely to be engaged in the educational activity. The odds that students are on-task compared to being off-task during direct instruction are about 19 to 1. On the other hand, in the absence of DI students are engaged about 57 percent of the time [$\hat{p}$(EN|NDI) = .573, S.D. = .142]. A further interpretation is that students are about 13 times more likely to be off-task (NE) when no DI is provided compared to being off-task when DI is provided.

The relation between DI and EN is asymmetrical, since the $\hat{p}$(EN|DI) is not equal to the $\hat{p}$(DI|EN): .967 $\neq$ .561 (the latter is not shown in Table 2). Furthermore, the standard error of the $\hat{p}$(EN|DI) is relatively small (.003). The 95 percent confidence interval for the $\hat{p}$(EN|DI) is .955 to .979. Since the sampling distribution is negatively skewed for a mean proportion this close to 1.0, a different method of determining the

Table 2. Results from Nonmetric Temporal Path Analysis: Proportion Time

System	DI*	EN	DI&EN	DI&NE	NDI&EN	NDI&NE	EN|DI	EN|NDI
1	.50	.80	.46	.04	.34	.16	.92	.67
2	.39	.49	.37	.02	.12	.49	.95	.20
3	.27	.56	.26	.01	.30	.43	.97	.41
4	.34	.69	.34	.00	.35	.31	1.00	.53
5	.48	.73	.47	.01	.25	.26	.98	.49
6	.40	.75	.39	.01	.35	.25	.98	.59
7	.44	.84	.40	.04	.44	.11	.91	.80
8	.36	.75	.33	.03	.42	.22	.92	.65
9	.30	.67	.29	.01	.39	.32	.96	.55
10	.32	.71	.31	.01	.40	.29	.98	.56
11	.42	.68	.42	.00	.26	.31	.99	.46
12	.38	.84	.37	.01	.47	.15	.97	.75
13	.31	.63	.31	.00	.32	.37	1.00	.46
14	.54	.87	.52	.02	.36	.11	.97	.77
15	.81	.92	.81	.00	.11	.08	1.00	.57
16	.67	.77	.62	.05	.15	.18	.93	.45
17	.24	.76	.24	.00	.52	.24	1.00	.69
18	.34	.74	.34	.00	.40	.25	.99	.61
19	.59	.87	.58	.01	.29	.12	.99	.71
20	.52	.64	.48	.04	.16	.33	.93	.33
21	.62	.83	.58	.04	.25	.13	.94	.66
22	.23	.65	.22	.01	.43	.34	.97	.56
23	.29	.79	.28	.01	.51	.20	.97	.71
24	.54	.75	.52	.02	.23	.24	.97	.49
25	.51	.82	.50	.00	.31	.18	.99	.63
Mean	.432	.741	.416	.015	.324	.243	.967	.573
S.D.	.144	.101	.139	.015	.114	.104	.029	.142

*Key:

DI = Direct Instruction
NDI = Non-direct Instruction
EN = Student Engagement
NE = Student Non-engagement

95 percent confidence interval is probably warranted. This is discussed further in a subsequent section.

A Comparison of NTPA and Correlational Analysis

NTPA and correlational analysis are mathematically related in certain special cases, although interpretation of the results differs due to assumptions about the nature of the relationship. The reason the two approaches are mathematically related in some instances is that the proportions or probabilities of categories in a classification are ipsative --i.e., they must sum to one by definition, since they are mutually exclusive and exhaustive of the classifcation. That is, for classification, *A*, with categories, $\underline{A}_1$ through $\underline{A}_n$:

$$p(A_1) + p(A_2) \ldots + p(A_n) = 1, \qquad p(A_i) \geq 0.$$

Furthermore, conditional probability is defined:

$$p(B_i | A_j) = p(B_i \ \& \ A_j)/p(A_j).$$

Thus:

$$p(B_i \ \& \ A_j) = p(B_i | A_j)p(A_j).$$

It should be noted that the formula for conditional probability will only be true empirically in NTPA when there is a one-to-one correspondence between occurrences in classifications *A* and *B*. Given the definition of NTPA and the time measure that is used here, these conditions are satisfied. Moreover, the relation, EN|DI, is not the same as the relation, 'If DI, then EN'. See Chapter 2. The latter implies a sequence, where DI must begin occurring <u>before</u> EN begins occurring. The former implies the <u>simultaneous</u> occurrence of EN and DI, given that DI is occurring--i.e., it does not matter whether EN or DI begins first.

Therefore, in the above example of the relation between EN and DI, the following equations are true, given the classical probability calculus:

$$p(EN) = p(EN \& DI) + p(EN \& NDI) \qquad [1]$$

$$p(DI) + p(NDI) = p(EN) + p(NE) = 1 \qquad [2]$$

$$p(EN|DI) = p(EN \& DI)/p(DI) \qquad [3]$$

$$p(EN|NDI) = p(EN \& NDI)/p(NDI) \qquad [4]$$

From [3] and [4]:

$$p(EN \& DI) = p(EN|DI)p(DI) \qquad [5]$$

$$p(EN \& NDI) = p(EN|NDI)p(NDI) \qquad [6]$$

From [2]:

$$p(NDI) = 1 - p(DI) \qquad [7]$$

From [6] and [7]:

$$p(EN \& NDI) = p(EN|NDI)(1 - p(DI)) \qquad [8]$$

From [5], [8] and [1]:

$$p(EN) = p(EN|DI)p(DI) + p(EN|NDI)(1 - p(DI))$$

$$= p(EN|DI)p(DI) + p(EN|NDI) - p(EN|NDI)p(DI) \qquad [9]$$

Regrouping:

$$p(EN) = p(EN|NDI) + [p(EN|DI) - p(EN|NDI)]p(DI) \qquad [10]$$

Equation [10] is of the form for a linear regression:

$$Y = A + BX \qquad [11]$$

If the conditional probabilities are known, then what appears to be a regression equation may be formed. From Table 1, $\hat{p}(EN|NDI) = .573$ and $\hat{p}(EN|DI) = .967$. Substituting these values into [10]:

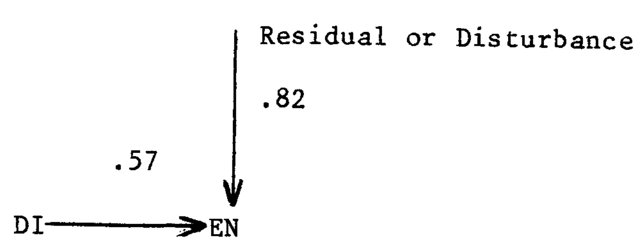
$$\begin{aligned} p(EN)' &= .573 + (.967 - .573)p(DI) \\ &= .573 + .394p(DI) \end{aligned} \qquad [12]$$

This is precisely the same regression equation that was obtained earlier in the correlational analysis with the exception of errors due to rounding. If the relation is viewed linearly and metrically, and if the proportional <u>amount</u> of DI is known, then the proportional <u>amount</u> of expected EN can be predicted from [12], although with little confidence due to the relatively small sample.

A similar equation predicting EN from NDI can be constructed:

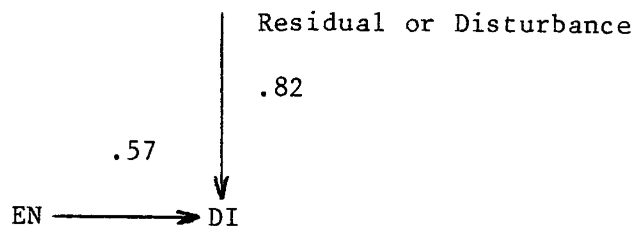
$$p(EN)' = .967 - .40p(NDI) \qquad [12']$$

If the raw regression weights are converted to standardized regression weights (or path coefficients), then the relationship of DI and EN can be depicted by a path diagram:



[P1]

In other words, an increase in DI by 1 S.D. unit is associated with and increase of .57 S.D. units of EN. Note that the path coefficient is -.57 if NDI is used to predict EN, since DI and NDI are ipsative and perfectly negatively correlated when there are only two categories in a classification. It is also the case that the following path diagram obtains:



[P2]

These identical path coefficients are due to the symmetrical nature of the correlation coefficient (equal to the standardized beta which is the path coefficient in the bivariate case). Both [P1] and [P2] are viable interpretations of the relationship. Adherents of metric path analysis would argue that theory should dictate which directionality is more plausible.

At this point one might be tempted to argue that correlational analysis is more adequate than NTPA since the former retains more information about the nature of the relation. That is, not only can the conditional probabilities be estimated at the points on the regression line where p(DI) equals zero and one, but also in the intermediate situations where p(DI) lies between zero and one. It may be argued that NTPA merely represents special cases of correlational analysis.

However, there is one assumption in NTPA that does not obtain in correlational analysis--i.e., a measure of a variable must be greater than or equal to zero and less than or equal to one. Only a limited set of regression equations can satisfy this assumption. For example, given the above derivation, what is the prediction of DI given knowledge of the proportion of EN? The following derivation is true according to the classical probability calculus (see [10]):

$$p(DI) = p(DI|EN) + [p(DI|EN) - p(DI|NE)]p(EN) \qquad [13]$$

Substituting known values:

$$p(DI)' = .058 + (.561 - .058)p(EN)$$

$$= .058 + .503p(EN) \qquad [14]$$

Thus, if EN is occurring, the p(DI) = .561 and if EN is not occurring, p(DI) = .058. However, the linear regression equation fitted to the sample data is:

$$p(DI)' = -.176 + .819p(EN) \qquad [15]$$

Making a similar interpretation, if EN is occurring (i.e., p(EN) = 1), the p(DI) = .643. But if EN is not occurring, the p(DI) = -.176. Both of these predictions are incorrect, given the known conditional probabilities from NTPA. Clearly, the NTPA and correlational analysis results differ in [14] and [15], respectively. The reason for the difference is that a probability or proportion is restricted to lie in the interval between zero and one inclusively. Beta weights and constants in regression analysis are not so constrained. According to the regression equation [15], the proportion of DI is predicted to be <u>negative</u> when the proportion of EN is zero. This result makes no sense. By definition a probability or proportion must be a positive real number between zero and one inclusively.

Upon more careful examination it appears that NTPA and correlational analysis will provide equivalent information only when it is the case that the slope of the regression line is such that when the independent variable is equal to zero or one, the predicted value of the dependent variable also lies between zero and one inclusively. This is possible only when certain features of the marginal distributions constrain the joint probability distribution.

In general, correlational analysis and NTPA will not be systematically related. This is due to the fact that there is usually no unique solution for obtaining joint probabilities given only marginal probabilities. Consider Table 3. If data are related in correlational analysis, only the marginal probabilities are known and can be utilized. Given the marginals, there is generally no unique solution for determining the joint probabilities. This can be proved as follows (all probabilities greater than or equal to zero and less than or equal to one):

$$p(DI) + p(NDI) = 1 \qquad [16]$$

$$p(EN) + p(NE) = 1 \qquad [17]$$

$$p(EN \,\&\, DI) + p(NE \,\&\, DI) = p(DI) \qquad [18]$$

$$p(EN \,\&\, NDI) + p(NE \,\&\, NDI) = p(NDI) \qquad [19]$$

$$p(EN \,\&\, DI) + p(EN \,\&\, NDI) = p(EN) \qquad [20]$$

$$p(NE \,\&\, DI) + p(NE \,\&\, NDI) = p(NE) \qquad [21]$$

If p(DI) and p(EN) are known, p(NDI) and p(NE) are uniquely determined by [16] and [17]. This is only true, however, when a classification

Table 3. Joint and Marginal Probabilities for the Simultaneous Occurrence of Types of Instruction and Student Orientation to Instruction. The Joint Probabilities Are in the Cells and the Marginal Probabilities Are the Respective Row and Column Totals of the Cells.

	EN	NE	
DI	A = p(EN & DI)	B = p(NE & DI)	K = p(DI)
NDI	C = p(EN & NDI)	D = p(NE & NDI)	L = p(NDI)
	M = p(EN)	N = p(NE)	

consists of <u>two</u> mutually exclusive and exhaustive categories. Thus, these values may be substituted into [18] through [21]. There are four equations and four unknown joint probabilities. To simplify notation, let A = p(EN & DI), ... etc., as given in Table 3. Re-writing [18] through [21] in matrix form and substituting letters for the probability expressions:

$$1A + 1B + 0C + 0D = K \quad (= 1 - L) \qquad [18']$$

$$0A + 0B + 1C + 1D = L \quad (= 1 - K) \qquad [19']$$

$$1A + 0B + 1C + 0D = M \quad (= 1 - N) \qquad [20']$$

$$0A + 1B + 0C + 1D = N \quad (= 1 - M) \qquad [21']$$

The determinate of the matrix of coefficients is zero:

$$\text{det.} = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{vmatrix} = 0$$

There is no unique solution to this set of equations, since the determinate of the matrix of coefficients is zero. This will be true in general, regardless of the number of categories in each classification and the number of classifications. The mathematical conclusion is that there is no way to uniquely determine the joint probability distribution given only the marginal probability distribution, except in a few special cases where the marginal probabilities are zeros and ones, or all equal.

On the other hand, the marginal probabilities do constrain the range of the joint probability distribution. Given the above equations, the following inequalities are derived:

$$M \geq A \leq K \qquad [22]$$

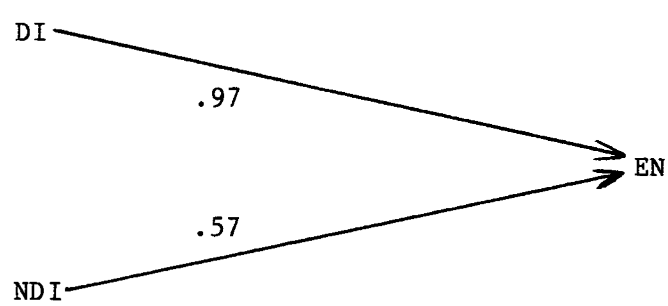
$$N \geq B \leq K \qquad [23]$$

$$L \geq C \leq M \qquad [24]$$

$$L \geq D \leq N \qquad [25]$$

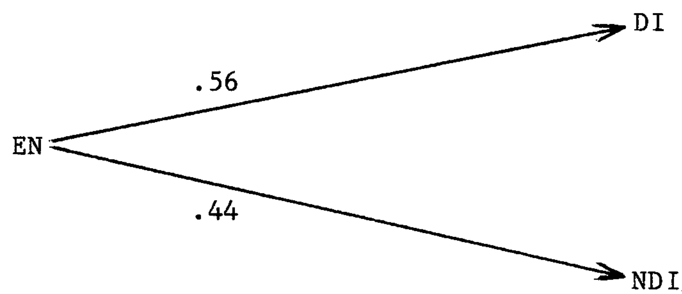
Moreover, even in the special cases when the conditional probabilities can be predicted from a linear regression equation, the variance of the conditional probabilities cannot be estimated, given only knowledge of the marginal probability distributions. In NTPA, not only can the conditional probabilities be estimated, but also their variances. Therefore, correlational analysis cannot be reduced to NTPA, although NTPA can be reduced to correlational analysis if a functional relation is desired.

The nonmetric path diagrams resulting from NTPA are as follows:



[P3]

and



[P4]

Compare [P3] and [P4] with [P1] and [P2]. Conventional metric path analysis provides two path coefficients of the same absolute value (±.57) for these four paths--i.e., directionality is undiscriminated due to the symmetric nature of the correlation coefficient, and the sign difference of the path coefficients is due to the ipsative nature of the categories in the two-category classification of type of instruction. In NTPA directionality is interpreted as conditional probabilities when the temporal sequence is unknown and as temporal conditional when the temporal sequence is known. In the sample data being investigated, temporal sequence is unknown, since the classifications were characterized conjointly at each sampling moment. Thus, if DI is occurring, the expected probability of EN also occurring at the same time is estimated to be .97. If NDI is occurring, the expected probability of EN also occurring at the same time is estimated to be .57. The variance of these mean probabilities can also be estimated and confidence intervals generated.

In metric path analysis, these conditional probabilities were not considered; hence, the path coefficients of the same absolute value were were obtained. In addition, when directionality (conditionality) is reversed in NTPA (see [P4]), different results are possible. This is due to the fact that a conditional relation is asymmetric. It is patent that NTPA provides more information about the nature of a relation than does correlational analysis. One can always perform a correlational analysis, if desired, given NTPA data; but one can do the converse only in a very restricted set of cases.

The Linear Models Approach: Analysis of Variance (ANOVA)

Cronbach (personal communication) asserted that NTPA merely represents an analysis of variance (ANOVA), where the design is sliced in an unconventional manner. Following his suggestion, a randomized block design can be generated, where subjects are treated as blocks and the independent variable is type of instruction. The dependent variable is student engagement (EN) and the levels of the independent variable are direct instruction (DI) and non-direct instruction (NDI). This is essentially a repeated measures design, where a measure of the proportion of EN is obtained for each subject (system) under each of the two so-called treatment conditions (DI and NDI). In Table 4 it can be seen that differences between the DI and NDI means are highly significant ($p < .001$). In standard interpretation the null hypothesis, that the difference between treatment means is equal to zero, is rejected. Thus, the alternative hypothesis that the difference between treatment means is greater than zero is not rejected (i.e., accepted). Strength of association between the independent and dependent variable is estimated to be .793, using the statistic, eta squared. The mathematical model upon which this ANOVA is predicated is given:

$$X_{ij} = \mu + \beta_j + \pi_i + \varepsilon_{ij}$$

Table 4. Analysis of Variance Summary of NTPA-Aggregated Data

Source	SS	df	MS	F
Type of Instruction	1.9416	1	1.9416	191.3***
Subjects (Systems)	.2602	24	0.0108	
Residual	.2437	24	0.0102	
Total	2.4455	49		

	p(EN|DI)	p(EN|NDI)
Mean	.967	.573
S.D.	.029	.142
Skewness	-.698	-.677
Kurtosis	-.702	+.796

*** $p < .001$

where:

μ = grand mean of treatment populations, which is constant for all observations.

β_j = effect of treatment j, which is constant for all observations within treatment population j, $\Sigma\ \beta_j = 0$.

π_i = a constant associated with block i, $\Sigma\ \pi_i = 0$.

ε_{ij} = experimental error, which is independent of other ε's and is normally distributed within each treatment population with mean = 0 and variance = σ^2_ε. (Kirk, 1968, p. 135)

If the assumptions which underlie this model are met, then ANOVA is a viable method of treating NTPA data aggregrated as conditional probabilities or proportions. These assumptions are now examined.

Independence of observations. In a traditional ANOVA it is assumed that observations are independent. That is, any given subject's score is presumed not to systematically influence any other subject's score on the dependent variable. This assumption appears to be met within each so-called treatment condition, since individual student-instructor systems were randomly sampled and each system represents a different classroom. However, there may be dependence of a subject's score in one treatment condition and his or her score in the other treatment condition. For example, a student who is more likely to be engaged during direct instruction may also be more likely to be engaged during no direct instruction. In this set of sample data the correlation between p(EN|DI) and p(EN|NDI) is equal to -0.0003. This is an additional assumption required for randomized block and other repeated measures designs. However, there is a direct test of this assumption, described below.

Observations drawn from normally distributed populations. Each of the treatment population distributions is assumed to be normal in form. Using the sample data as estimates of the respective population distributions, both distributions are negatively skewed; kurtosis is negative for p(EN|DI) and positive for p(EN|NDI). See Table 4. Statisticians have shown that ANOVA is fairly robust with respect to violations of normality if the sample size is relatively large and the treatment groups are of equal size (Kirk, 1968; Hays, 1973). Moreover, it is recommended that a suitable transformation be used on the data to make the distribution more normal in form when this assumption is violated. For proportional data, as is the case here, an arcsin transformation is often recommended, particularly since proportions close to zero or one necessarily have skewed sampling distributions.

Homoscedasticity. Variances of the population treatment groups are assumed to be equal. This assumption appears to be violated seriously with these sample data. F_{max} (Hartley, 1950) is 24.86 and statistically significant (df = 2,24, $p < .05$), thus indicating the variances are not homogeneous. Cochran's C (1941) is .961, also statistically significant (df = 2,24, $p < .05$).

Homogeneity of population covariances. An additional assumption is necessary in a randomized block design. It is assumed that the population covariances for all pairs of treatment levels are homogeneous, as well as the variances within treatment levels. The symmetry of the variance-covariance matrix can be tested for significance by means of a procedure described by Kirk (1968). The observed chi square is 41.874, which is greater than $\chi^2_{.05,\ df=1} = 3.841$, implying that the data depart from the required symmetry. This should be expected, since the ances are equal in the two-treatment situation and it was already concluded that the variances were not homogeneous.

These assumptions are required in order to make a valid inference from a univariate F test. If a multivariate test such as Hotelling's T^2 is used, the only assumption is that the populations are multivariate normal.

Comparison of NTPA and ANOVA

There is a more basic issue, however, even if the statistical assumptions of ANOVA are met in NTPA. Ordinarily in ANOVA, the score for each case is considered to be the outcome of a trial--i.e., the result of manipulation of a level of the independent variable. These trials are replicated across cases which are randomly selected and assigned to treatment conditions. In NTPA the method of aggregation is atypical for ANOVA. That is, a score for each case (system) in a treatment condition represents the proportion of outcomes which were EN when the so-called treatment was present. There was no true experimental manipulation of treatments (types of instruction, either DI or NDI). Rather, so-called treatment conditions were observed as they naturally occurred in each classroom. Some students were exposed to more or less of a given treatment condition than others. By converting to proportions, the amount of each treatment condition was normalized. This is clearly an atypical manner of deriving measurements of each case under each treatment condition.

More importantly, in ANOVA it is assumed that treatment effects are linear and additive, as represented by the general mathematical model:

$$X_{ij} = \mu + \beta_j + \varepsilon_{ij} .$$

Each case's score is assumed to consist of the grand mean, plus a deviation from the grand mean (β_j), plus an error component. The treatment effect (β_j) for each case is assumed to be constant for all cases within a treatment condition and errors are assumed to be normally and independently distributed. Moreover, errors are presumed to be uncorrelated with treatment conditions and with each other.

In NTPA no assumptions are made concerning linearity and additivity in the relationship between independent and dependent variables, nor is any distinction made between independent and dependent variables. Rather, the nonmetric temporal path itself is considered in NTPA as a single random variable. The connection between or among the parts of a relation is viewed as temporal (i.e., sequential, simultaneous or conditional). A prior or conditional part of a relation is not considered as a treatment in the experimental sense. Thus, temporal relations are simply observed, enumerated and converted to proportions (or relative frequencies) for each system under consideration. The results of NTPA do not explain why a relation occurs, but simply that it does with an estimated probability.

General Remarks Concerning NTPA and the LMA

It has been demonstrated that NTPA data can also be subjected to the linear models approach (LMA--correlational analysis and ANOVA), providing that one is willing to make additional assumptions about the nature of the relationship among factors (i.e., that it is deterministic) and that those assumptions are fulfilled. However, the converse does not obtain. Although the LMA is derivable from NTPA, NTPA is not derivable from the LMA. This is due to the fact that a deterministic system is a special case of a stochastic system.

Concerning statistical inference in NTPA, the only assumption necessary is that of independence of observations. Replications of occurrences of a relationship <u>within</u> a system are not considered to be independent, although different systems are considered to be independent. For example, this procedure is no different than summing a respondents's item scores on an achievement test and forming a proportion indicating the percent of correct answers. A person's responses to different items on the test are not likely to be independent. This causes statisticians no problems as long as one person's responses do not affect another's responses on the test.

If systems are randomly sampled from a population, then conventional inferential statistics may be employed (i.e., with the unit of analysis as the system). There is a problem, however, of not knowing which theoretical sampling distribution is appropriate for estimating a confidence interval which is likely to include the population mean for a given nonmetric temporal path. As mentioned earlier, values of a proportion which are close to zero or one will have skewed sampling distributions. The problem of which theoretical sampling distribution is appropriate becomes less significant when sample sizes become relatively large ($\underline{n}$ > 500), since the central portions of the various theoretical distributions (e.g., binomial, beta, Poisson, Gaussian, chi square, etc.) are very similar in form (Schmidt, 1969).

An alternative approach to estimating a confidence interval is Tchebycheff's inequality:

$$prob\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{1}{k^2}\ .$$

The probability that a standard score drawn at random from the distribution has absolute magnitude greater than or equal to some positive number k is always less than or equal to $1/k^2$, regardless of the mathematical form of the distribution (k is the number of sigma units) (Hays, 1973). For example, given a distribution with some mean and variance, the probability of drawing a case having a standardized score of 4 or more must be at most $1/4^2$, or 0.0625. Tchebycheff's inequality is clearly a more conservative approach than one where the form of the sampling distribution is known.

If it is assumed that the distribution of the random variable is both symmetric and unimodal, then the inequality becomes (Hays, 1973):

$$prob\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{4}{9k^2}\ .$$

If this assumption is made, then the probability of randomly drawing a case having a z-score of 2 or more must be at most $4/[9(2^2)] = 4/36 =$ 0.11.

And finally, if it is assumed that the distribution is normal in form, then the probability of obtaining a z-score of 2 or more is about 0.046.

One might argue that the binomial or chi square distribution would be appropriate for NTPA. Use of these theoretical distributions does not appear to be legitimate in NTPA. To use the binomial distrbution one must assume a stationary Bernoulli process--i.e., that each trial is independent and has the same probability of an outcome as any other trial. If each instance of a relation (i.e., trial) in a system is independent of each other instance, then these distributions could be legitimately used. However, this assumption appears unwarranted in NTPA. For example, student engagement (EN) on one occasion may be related to that same student's engagement on another--i.e., trials (replications) within a system are not likely to be independent. The same argument obtains for not using chi square as a theoretical sampling distribution.

Summary

It has been demonstrated by an empirical example and proved mathematically that the LMA cannot be reduced to NTPA, due to the difference in approach to the measurement of a relation. Additional assumptions are made in the LMA which are not made in NTPA. It was further shown that if these assumptions are made (and fulfilled) NTPA-aggregated data can be analyzed by the LMA. However, the LMA conclusions are restricted by the assumptions of linearity and additivity.

In making statistical inferences about population parameters of an NTPA relation, it is the case that the form of the mathematical sampling distribution is unknown, thus making it more difficult to legitimately construct a confidence interval that is likely to contain the population mean. Two alternatives were suggested: use large samples or Tchebycheff's inequality. A third alternative is to transform probabilistic measures of a relation by the arcsin transformation or some other suitable transformation so as to approximate a normal distribution. Then one can more legitimately construct confidence intervals around transformed score means and subsequently convert these back to the original metric in a mannner similar to Fisher's _r_ to _z_ transformation.

Data collected by means of NTPA contain more information than in the LMA. Thus, it is possible to investigate relations _a posteriori_ in the data that may not have been anticipated when the data were initially collected, since the raw NTPA data preserve the temporal nature of occurrences of events characterized by the classifications of interest. Theoretically, given the _a priori_ classifications, all possible nonmetric temporal paths could be investigated, although the number of possible temporal paths expands geometrically with the number of classifications, categories within each classification, and the lengths of paths investigated. Theory is expected to guide the selection and investigation of relations in NTPA, as it is for scientific inquiry in general.