American Educational Research Journal
Spring 1990, Vol. 27, No. 1, pp. 180–204

Analysis of Patterns in Time: A Method of Recording and Quantifying Temporal Relations in Education

Theodore W. Frick
Indiana University

Analysis of patterns in time (APT) is a method for gathering information about observable phenomena such that probabilities of temporal patterns of events can be estimated empirically. If appropriate sampling strategies are employed, temporal patterns can be predicted from APT results. As an example of the fruitfulness of APT, it was discovered in a classroom observational study that elementary students were on task 97% of the time if some form of direct instruction was occurring also, whereas they were on task only 57% of the time during nondirect instruction. As a second example, APT results were used as a rule base for an expert system in adaptive computer-based testing. When two different computer tests were studied, average samples of 9 and 13 test items were required to make mastery and nonmastery decisions when items were selected at random. These decisions were, respectively, 94% and 98% accurate compared to those reached from two much larger test item pools. Finally, APT is compared to the linear models approach and event history analysis. The major difference is that in APT there is no mathematical model assumed to characterize relations among variables. In APT the model is the temporal pattern being investigated.

THEODORE W. FRICK is an associate professor in the Department of Instructional Systems Technology, School of Education, Indiana University, Bloomington, IN 47405. His specializations are educational inquiry methodology and computer-assisted instruction.

Educational measurement and statistical analysis of results historically have tended to follow a pattern where variables are measured separately and then a mathematical model is chosen to portray the relationship among the variables. Most often a linear models approach (LMA) has been adopted, such as analysis of variance, multiple regression analysis, discriminant analysis, path analysis, time series analysis, and so on. The LMA can be characterized as follows: Measure variables separately, then relate them mathematically (Frick, 1983).

In the past two decades alternative research methodologies have gained attention in the educational research community (cf. Guba & Lincoln, 1981; Maccia & Maccia, 1976). One such method of collecting and analyzing evidence to help answer educational research questions is analysis of patterns in time (APT).[1] This alternative view might be characterized as follows: Measure temporal relations directly by counting their occurrences (Frick, 1983). APT requires a subtle but significant shift in one's world view, compared to that often taught in educational measurement and statistics courses.

An educational researcher with an APT world view is comparable to an epidemiologist who observes that middle-aged persons who take a small dosage of aspirin daily are subsequently less likely to suffer a heart attack than those who do not. Another example of an APT world view is a baseball manager who observes how often each player has hit safely against left-handed pitchers when runners are in scoring position. In both cases temporal patterns are observed and enumerated rather than estimating beta weights for regression analysis or means for ANOVA, as could be done in the LMA.

Knowledge of likelihoods of temporal patterns can be used to predict subsequent events and aid decision makers, for example, for forecasting. Although temporal patterns do not necessarily indicate causal relationships, such patterns may provide good leads to further experimental research.

I have formalized analysis of patterns in time through adoption of fundamental concepts from information theory, set theory, and probability theory. This formalization was also influenced by the SIGGS Theory Model developed by Maccia and Maccia (1966). Before describing this formalization, I will present an example of the fruitfulness of an APT view. The explication of APT is followed by another example and a discussion in which major extant methodologies are compared to APT.

An Example of APT Results From a Classroom Observational Study

I originally conceived of APT in the mid-1970s as a methodology of classroom observational research to investigate patterns of transactions among students, teachers, curricula, and educational settings. As predicta-

ble patterns are discovered, these can contribute to pedagogical knowledge and perhaps lead to better pedagogical theory.

APT principles were applied to the design of a classroom observation system for investigating academic learning time of handicapped students (Frick & Rieth, 1981). In this system numerous classifications were used, including types of instructional groupings, student task success, subject matter, types of instructional activities, student task engagement, and types of instructor behaviors (questions, feedback, explanations, etc.).

Observational data were collected on 25 mildly mentally handicapped students and their teachers as part of a study of academic learning time and student achievement (Rieth & Frick, 1982). Students were observed a total of 8 to 10 hours each at different times during the school day over a period of about 6 months. Using the Academic Learning Time Observation System (ALTOS) (Frick & Rieth, 1981), highly trained observers collected observational data on paper-and-pencil coding forms. During mathematics and language arts activities observers coded target student and instructor behaviors at 1-minute intervals.

For illustration, only two classifications (with categories in parentheses) are discussed: available instruction (direct, nondirect, null), and student orientation to academic instruction (engaged, nonengaged, null).

The kind of available instruction was viewed from the point of the target student: *Direct* instruction was defined as 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 target student, the source of direct instruction could be the teacher, another person in the class, such as a peer or an aide, or something capable of sending information to and receiving information from the student (e.g., a computer-based instructional program). If there was no academic transaction with the target student or group containing the student during an academic educational activity, then the type of available instruction was considered to be *nondirect*. If no academic educational activity was occurring, then available instruction was coded as *null*.

The observers also coded the type of target student orientation to academic instruction that was occurring simultaneously with the type of available instruction. The target student was considered to be *engaged* in an academic activity if she or he appeared to be attending to the substance of that activity. If the student clearly was not attending to the academic substance (e.g., off-task behavior), he or she was coded as *nonengaged*. If the student was participating in a nonacademic activity, *null* was coded.

Due to insufficient funds and lack of truly portable computers at the time of the study (1981 – 1983), paper-and-pencil coding forms were used, and point-time sampling of classroom events was done at 1-minute intervals. Therefore, true sequential patterns of interaction could not be quan-

tified. APT time measure functions, however, could be applied to nearly 15,000 one-minute samples collected on all 25 students. (See Frick, 1983, 1988, for counting rules for time measure functions, frequency measure functions, joint occurrence, sequential occurrence, etc.)

Results of these time measure functions for the 25 systems are presented in Table 1. Each system consisted of the target student and his or her classroom environment(s) and teacher(s). Some target students spent time in a special or resource classroom and in a regular elementary

Table 1
Results From APT Time Measure Functions in the Academic Learning Time Study

	Proportion of time							
S	DI	EN	DI ∩ EN	DI ∩ NE	ND ∩ EN	ND ∩ NE	EN|DI	EN|ND
1	0.50	0.80	0.46	0.04	0.34	0.16	0.92	0.67
2	0.39	0.49	0.37	0.02	0.12	0.49	0.95	0.20
3	0.27	0.56	0.26	0.01	0.30	0.43	0.97	0.41
4	0.34	0.69	0.34	0.00	0.35	0.31	1.00	0.53
5	0.48	0.73	0.47	0.01	0.25	0.26	0.98	0.49
6	0.40	0.75	0.39	0.01	0.35	0.25	0.98	0.59
7	0.44	0.84	0.40	0.04	0.44	0.11	0.91	0.80
8	0.36	0.75	0.33	0.03	0.42	0.22	0.92	0.65
9	0.30	0.67	0.29	0.01	0.39	0.32	0.96	0.55
10	0.32	0.71	0.31	0.01	0.40	0.29	0.98	0.56
11	0.42	0.68	0.42	0.00	0.26	0.31	0.99	0.46
12	0.38	0.84	0.37	0.01	0.47	0.15	0.97	0.75
13	0.31	0.63	0.31	0.00	0.32	0.37	1.00	0.46
14	0.54	0.87	0.52	0.02	0.36	0.11	0.97	0.77
15	0.81	0.92	0.81	0.00	0.11	0.08	1.00	0.57
16	0.67	0.77	0.62	0.05	0.15	0.18	0.93	0.45
17	0.24	0.76	0.24	0.00	0.52	0.24	1.00	0.69
18	0.34	0.74	0.34	0.00	0.40	0.25	0.99	0.61
19	0.59	0.87	0.58	0.01	0.29	0.12	0.99	0.71
20	0.52	0.64	0.48	0.04	0.16	0.33	0.93	0.33
21	0.62	0.83	0.58	0.04	0.25	0.13	0.94	0.66
22	0.23	0.65	0.22	0.01	0.43	0.34	0.97	0.56
23	0.29	0.79	0.28	0.01	0.51	0.20	0.97	0.71
24	0.54	0.75	0.52	0.02	0.23	0.24	0.97	0.49
25	0.51	0.82	0.50	0.00	0.31	0.18	0.99	0.63
Mean	0.432	0.741	0.416	0.015	0.324	0.243	0.967	0.573
SD	0.144	0.101	0.139	0.015	0.114	0.104	0.029	0.142

Note. S = system; DI = direct instruction; EN = student engagement; ND = nondirect instruction; NE = student nonengagement.

classroom, but this was taken as one system for each target student.) For each system, data were aggregated using APT time measure functions. For example, in system 1, direct instruction was made available to the target student 50% of the time. That student was engaged 80% of the time overall. The joint occurrence of direct instruction and student engagement occurred 46% of the time, and so on. The proportion of engagement, given that direct instruction was occurring at the same time was 0.92 for that student, whereas the student was engaged only 67% of the time during nondirect instruction.

Perhaps the most important finding across the 25 systems was the very high proportion of student engagement during direct instruction (0.967), compared to engagement during nondirect instruction (0.573). In other words, students were about 13 times more likely to be off task during nondirect instruction than during direct instruction.

The linear correlation between direct instruction and student engagement was about 0.57, and although significant at the 0.05 level, the linear model does not reveal the clear pattern indicated by the APT time measure functions for joint events. Although we cannot infer that direct instruction causes high student engagement, we can nonetheless predict that, if direct instruction is occurring, the probability of mildly handicapped student engagement in the elementary grades is extremely high. This relationship was not anticipated before the study. The pattern was consistent across different kinds of settings and subject matter areas. What caught the investigators' eyes was the consistently low amount of joint occurrences of direct instruction and student nonengagement compared to nonengagement and nondirect instruction.

Analysis of Patterns in Time

APT is based on set theory, information theory, and probability theory (cf. Frick, 1983, 1988; Maccia & Maccia, 1966). In set theory a *relation* is taken as a subset of the Cartesian product of two or more sets of elements. A relation is thus a set of ordered pairs if two sets are in the Cartesian product, or more generally a set of n-tuples if more than two sets are involved. Each n-tuple symbolizes the specific joining of elements. In information theory, categories in a classification are analogous to elements in a set with the added condition that categories are mutually exclusive and exhaustive. By taking *information* as a characterization of occurrences, observed events or states of affairs can be mapped into categories in classifications (Maccia & Maccia, 1966).

An H measure traditionally has been used in information theory to indicate the uncertainty in the probability distribution derived from observed occurrences of categories in a single classification (cf. Coombs, Dawes, & Tversky, 1970). Uncertainty is zero when the probability of a particular category is one and all other categories have zero probabilities. Uncer-

tainty is maximum when each category is equally likely to occur.

A T measure in information theory has been used to indicate transmission of information. A T measure can be determined from the joint probability distribution derived from observations of paired event occurrences characterized by the Cartesian product of the classifications.

When time of occurrence of observed events is considered as well, such a mapping represents a temporal pattern for each of the n-tuples of categories. Maccia & Maccia (1966) recommended T as a measure of system feed in and feed out, that is, for transmission of information from a negasystem to a system and vice versa. It is patent that H and T are inappropriate for measuring uncertainty of particular temporal patterns represented by each of the n-tuples. Instead, the probability of occurrence of the nth element of a temporal pattern (n-tuple) can be estimated by simply observing how often it occurs following occurrences of preceding elements. Although H and T measures can be determined also, they are not essential to APT.

To illustrate basic APT concepts and procedures, an example from observation of the weather is discussed first. In Figure 1 an APT *score* derived from observation of the weather is presented. This score is analogous to the notation by someone who listens to an orchestra playing and writes the different parts as they are heard, using traditional staffs and musical notation (i.e., recreates the musical score).

The classifications used in Figure 1 are those that an amateur meteorologist might use to characterize occurrences of weather: cloud structure, precipitation, atmospheric pressure, air temperature, and season of year. Each classification consists of a set of mutually exclusive and exhaustive categories. At any point in time, one and only one category may be used to characterize the current state of a classification in APT. For example, if the season of year is categoried as *spring* at some point, then it cannot be winter, summer, or fall at the same time.

Classifications for the observation system used in Figure 1 and their categories (in parentheses) are as follows: cloud structure (cumulus, nimbus stratus, nimbus cumulus, cirrus, null); precipitation (rain, sleet, snow, null); atmospheric pressure (above 30, below 30 (p.s.i),, null); air temperature ($-50°$F, $-49°$F, . . . , 119°F, 120°F, null); and season of year (winter, spring, summer, fall, null).

The task of an observer who is creating an APT score is to characterize simultaneously the state of each classification as events relevant to the classifications change over time. To help simplify the task for an observer, he or she only need note when there is a category change in a classification and the time the change occurs—for example, the precipitation changes from rain to sleet at 9:38:18 a.m., or the cloud structure overhead changes from nimbus stratus to null at 12:23:38 p.m.

When the observation begins, the state of each classification is noted

Categorization of Event Changes

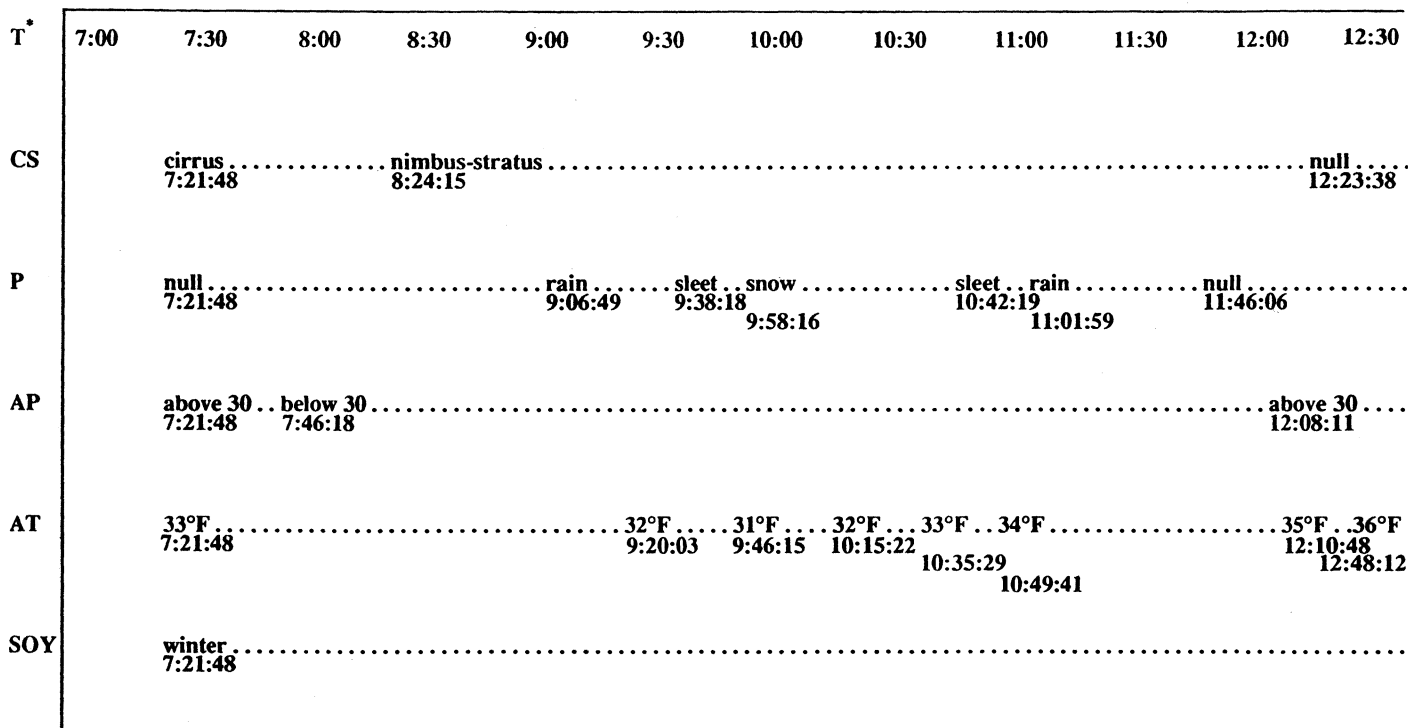

Figure 1. **An APT score resulting from observation of the weather**

Note. *T: Time (HH:MM); CS: Cloud Structure; P: Precipitation; AP: Atmospheric Pressure; AT: Air Temperature; SOY: Season of Year.

along with the time. As can be seen in Figure 1, at 7:21:48 a.m. the cloud structure is cirrus, precipitation is null, atmospheric pressure is above 30 (p.s.i.), air temperature is 33°F, and the season is winter. The observer waits until there is a change in one or more classifications before recording further. For example, in Figure 1 the first change observed was at 7:46:18 a.m. when the atmospheric pressure dropped to below 30 (p.s.i.). The next change was at 8:24:15 a.m. when the cloud structure directly overhead changed to nimbus stratus. The precipitation changed to rain at 9:06:49 a.m., the air temperature dropped to 32°F at 9:20:03 a.m., the precipitation changed to sleet at 9:38:18 a.m., and so on.

Notice that during the observation the season of year never changed. After recording that the season of year was winter at the beginning of the observation, no further changes were recorded in that classification. Also notice that each classification has a null category, meaning that there is nothing occurring that is relevant to the classification to characterize at that point in time.

By scanning across an APT score, the sequence of changes within and among classifications can be seen. If one scans vertically at some point in the APT score, the joint occurrence of categories in different classifications can be observed. For example, at 10:30 a.m. the state of affairs is that cloud structure is nimbus stratus, precipitation is snow, atmospheric pressure is below 30, air temperature is 32°F, and season of year is winter. The joint occurrence of categories from different classifications is analogous to musical harmony, and more specifically to the variety of orchestral sounds that occur when the various timbres of different groups of simultaneously played instruments are combined.

An APT score is an observational record. The notion of a score in APT is different from the conventional usage of *score*, such as the score in a soccer match or a student's score on an achievement test. In APT a score is the temporal configuration of observed events characterized by categories in classifications (e.g., see Figure 1). The notion of a score in APT is akin to that of a musical score. In APT it is as if someone like Mozart heard a group of musicians playing and was able to write out the score as he was listening.

Making Queries About APT Scores

We could be satisfied by studying APT scores, looking for recurring patterns or combinations of events, and making further note of them. If we are interested in quantification, we can count, for example, how many times a particular consequent event follows a particular antecedent event; or we can aggregate durations of certain kinds of events to see what proportion of the overall time they occupy. For example, we might ask, How often or what proportion of the time is it the case that

1. precipitation is rain?
2. if atmospheric pressure is above 30,
   then precipitation is rain or sleet or snow?
3. if cloud structure is nimbus stratus
      and atmospheric pressure is below 30,
   then precipitation is sleet or snow?
4. if air temperature is 32°F or 33°F or 31°F
      and atmospheric pressure is below 30
      and season of year is winter,
   then precipitation is sleet or snow?
5. if atmospheric pressure is above 30,
   then atmospheric pressure is below 30
      and cloud structure is nimbus-stratus,
   then precipitation is not null?

In APT, questions such as these are referred to as *queries*. Given a query, an APT score is scanned such that instances and durations of the specified pattern are aggregated. For example, if these queries are made about information in the APT score illustrated in Figure 1, the results displayed in Table 2 obtain.

Notice that the results of APT queries are given according to phrases in each query. Query 1 has one phrase; 2, 3, and 4 have two phrases; 5 has three phrases. Each phrase is terminated with a comma, and the last phrase of a query ends with a question mark.

Aggregate data are reported by APT query phrase. For example, the first and only phrase in query 1 was found in the score to be true 2 out of 5 times. This means that the precipitation changed to rain twice out of the 5 recorded instances of precipitation changes. Given these data, the likelihood of a change to rain is 2/5, or 0.40. The total duration in which this phrase was true was 4,536 seconds out of the 19,584 seconds of observation (length of the APT score). Rain occurred 23.2% of the time (4,536 × 100/19,584). Note that null codes are not counted when determining frequency of changes in a classification, but are taken into consideration for duration of categories. For example, the null code in the precipitation classification at 11:46:06 a.m. indicates that the rain that began at 11:01:59 has stopped.

Query 2 is a two-phrase query. The first phrase, "atmospheric pressure is above 30," was found to be true in the APT score in two out of three changes. If a query phrase in APT begins with the key word *then*, it can be true in the data only if the preceding phrase has first become true. No instances of the second phrase, "then precipitation is rain or sleet or snow?" were found to be true in the data, given that the first phrase was true. Although the sample is relatively small to make any generalizations here, if the pattern specified by query 2 obtained across many samples, then

Table 2
Results of APT Queries About the Score Illustrated in Figure 1

Query	Frequency	Likelihood	Time (in seconds)	% time
(a) Precipitation is rain?				
	2 out of 5	0.40000	4,536 out of 19,584	23.16177
(b) If atmospheric pressure is above 30,				
	2 out of 3	0.66667		
then precipitation is rain or sleet or snow?				
	0 out of 0	No data	0 out of 19,584	0.00000
(c) If cloud structure is nimbus stratus				
and atmospheric pressure is below 30,				
	1 out of 5	0.20000		
then precipitation is sleet or snow?				
	3 out of 5	0.60000	5,021 out of 19,584	25.63828
(d) If air temperature is 32°F or 33°F or 31°F				
and atmospheric pressure is below 30				
and season of year is winter,				
	5 out of 10	0.50000		
then precipitation is sleet or snow?				
	3 out of 4	0.75000	1,945 out of 19,584	9.93158
(e) If atmospheric pressure is above 30,				
	2 out of 3	0.66667		
then atmospheric pressure is below 30				
and cloud structure is nimbus stratus,				
	1 out of 3	0.33333		
then precipitation is not null?				
	5 out of 5	1.00000	9,557 out of 19,584	48.80004

these results could be used to make predictions about the weather. That is, if the barometric pressure is above 30 pounds per square inch, the likelihood of subsequent rain, sleet, or snow would be very low.

Query 3 also has two phrases. The first phrase contains two phrase segments connected by the conjunction *and*. For a multisegment phrase to become true in the data, all segments must be found to be true, although the temporal order in which these segments occur is irrelevant. Only when all segments in a phrase have become true in the data do we look for instances of the next phrase. Given the data in Figure 1, 3 out of 5 changes in precipitation were either sleet or snow, after it was first true that the cloud structure was nimbus stratus and it was true that the atmospheric pressure was below 30.

Query 4 has two phrases, and the first indicates an even more complex antecedent condition must obtain. Query 5 illustrates a three-phrase

query. First, it must be true in the data that atmospheric pressure is above 30. Next, it must be true that atmospheric pressure changes to below 30 and cloud structure becomes nimbus stratus (or vice versa). Only when the first two phrases have become true, in the order specified, do we look for instances of the third phrase. It so happens in these data that precipitation was not null in five out of five changes. If this pattern were to obtain across numerous samples of weather observation, then we would predict that some kind of precipitation is highly likely following a change in barometric pressure from above 30 p.s.i. to below 30 p.s.i. in combination with the appearance of nimbus stratus clouds. Of course, we would be extremely cautious here, due to a very small sampling of weather occurrences.

See Frick (1983, 1988) for further information on query syntax, counting rules, and computer software logic to aid in the process of data collection and analysis. The brief examples just described illustrate most of the basic features of APT, but do not deal with complexities of longer queries or the issue of recursive queries in which different phrases indicate the same classifications.

A Second Example: APT and Adaptive Computer-Based Testing

Frick, Plew, and Luk (1989) have been researching an expert systems approach to computer-based testing that builds on APT conceptions. In essence, the results of APT queries are stored in a computer data base as expert systems rules with associated probabilities. This information is used by a computer-based testing system to make mastery and nonmastery decisions about students taking a test.

The most prevalent extant method of adaptive testing is based on item response theory (IRT) (cf. Lord & Novick, 1968; Weiss & Kingsbury, 1984). This method has been shown to be effective in estimating examinee ability with selection of a subset of items that match his or her ability level (Weiss & Kingsbury, 1984). One limitation of IRT-based adaptive testing is that a very large sample of examinees (from 200 to 1,000, depending on whether the one-, two-, or three-parameter mdoel is chosen) must be tested in advance to obtain reasonably accurate estimates of item parameters used in the decision functions.

Clearly, the IRT approach is not practical unless one has access to large samples of examinees as do many testing bureaus. Frick, Plew, and Luk (1989) have invented an alternative approach, termed EXSPRT, which apparently requires many fewer examinees (a minimum of 50) to develop an initial APT-generated rule base.

An Example of EXSPRT

Suppose that we have developed a pool of test items that matches a par-

ticular instructional objective and that our goal is to decide whether or not a particular student has mastered that objective (e.g., Mager, 1973). Suppose further that our aim is to administer no more questions than are necessary to reach a mastery or nonmastery decision, and yet we want to be highly confident in our decision.

First, we need to construct a rule base. There are various ways that this could be done, but let us use a straightforward empirical approach. We obtain a sample of students representative of those who would be likely to be learning the instructional objective, who are learning, and who have learned (e.g., third-grade students and multiplication of two-digit numbers; college freshmen taking a course in probability theory; graduate students in education learning how computers work).

Next, we give the whole test to this sample of students. We must then decide on a cutoff score for determining mastery and nonmastery. Suppose we are satisfied that anyone who scores 85% or higher on the test has minimally mastered the instructional objective being tested. This allows us to sort students into a mastery group and a nonmastery group. We now construct a rule set for each test item. For example (these are fictitious data, used for illustration only):

Rule 1.1. If achievement status is mastery and item number is 1, then student answer is correct: APT probability = 0.92.

Rule 1.2. If achievement status is mastery and item number is 1, then student answer is incorrect: APT probability = 0.08.

Rule 1.3. If achievement status is nonmastery and item number is 1, then student answer is correct: APT probability = 0.47.

Rule 1.4 If achievement status is nonmastery and item number is 1, then student answer is incorrect: APT probability = 0.53.

A quadruplet of such rules can be constructed for each item on the test, based on the proportions of masters and nonmasters, respectively, who answered the item correctly and incorrectly. We will assume that our student sample is large enough and representative enough of the population of those students of interest that we have sufficient confidence in the data used to derive the rules. The rules can be more conveniently summarized in tabular format. Some hypothetical data are provided below:

Item	$P(m \cap q \rightarrow c)$	$P(m \cap q \rightarrow i)$	$P(n \cap q \rightarrow c)$	$P(n \cap q \rightarrow i)$
1	0.92	0.08	0.47	0.53
23	0.81	0.19	0.24	0.76
38	0.98	0.02	0.86	0.14
63	0.89	0.11	0.65	0.35

where m = mastery, n = nonmastery, q = question, c = correct, and i = incorrect.

Now, we will use the rule base to decide the mastery status of a particular student about whom we presently know nothing with respect to mastery or nonmastery of the instructional objective assessed by the test items. Therefore, our prior probabilities of mastery and nonmastery are equal to 0.50 for this student.

Observation 1. We randomly select an item from the pool (#63). We administer it to this student, who answers it incorrectly. Our expert systems inference engine will reason according to Bayes' theorem as follows (cf. Schmitt, 1969):

Alternative	Prior probability of alternative		Probability of Alternative and #63→ i		Joint probability	Posterior probability
Mastery	0.50	×	0.11	=	0.055 ÷ Sum	= 0.239
Nonmastery	0.50	×	0.35	=	0.175 ÷ Sum	= 0.761
			Sum	=	0.230	

The prior probability of each alternative is multiplied by the probability of the observation, given that the alternative is true. The estimated APT probability of an incorrect response by a master for item #63 is 0.11, which when multiplied by 0.50 yields a joint probability of 0.055. Similarly, the estimated APT probability of an incorrect response by a nonmaster for item #63 is 0.35, and when multiplied by the prior probability of nonmastery (0.50), results in a joint probability of 0.175. The joint probabilities are normalized by dividing each by the sum of the joint probabilities. After this observation, the posterior probability for mastery is now 0.055 ÷ 0.23 = 0.239. The posterior probability for the nonmastery alternative is 0.175 ÷ 0.23 = 0.761. At this point, the nonmastery alternative is about three times more likely than the mastery alternative.

Observation 2. We continue testing by selecting another item at random from the pool. We give item #23 to the student, who answers it correctly. We update as follows, only this time we use the most recent posterior probabilities as our new priors:

Alternative	Prior probability of alternative		Probability of Alternative and #23→ c		Joint probability	Posterior probability
Mastery	0.239	×	0.81	=	0.194 ÷ Sum	= 0.515
Nonmastery	0.761	×	0.24	=	0.183 ÷ Sum	= 0.485
			Sum	=	0.377	

This time in the third column we use the probability of a correct response to item #23, given each alternative. The odds of nonmastery to mastery have now become about equal, given the two observations made thus far.

Observation 3. This time we select at random item #1, which the student answers incorrectly. We update, as before, using the most recent posterior probabilities as our new priors.

Alternative	Prior probability of alternative		Probability of Alternative and #01→ i		Joint probability	Posterior probability
Mastery	0.515	×	0.08	=	0.041 ÷ Sum	= 0.138
Nonmastery	0.485	×	0.53	=	0.257 ÷ Sum	= 0.862
			Sum	=	0.298	

The odds are a little over 6 to 1 in favor of nonmastery at this point.

Observation 4. We select another item, #38, at random, which our student also misses.

Alternative	Prior probability of alternative		Probability of Alternative and #38→ i		Joint probability	Posterior probability
Mastery	0.138	×	0.02	=	0.003 ÷ Sum	= 0.024
Nonmastery	0.862	×	0.14	=	0.121 ÷ Sum	= 0.976
			Sum	=	0.124	

After the fourth observation, the posterior probability of the nonmastery alternative is about 0.98, roughly 40 times as great as the probability that the mastery alternative is true. Should we stop the test now? If so, on what basis? It appears that it is extremely likely that this particular student is a nonmaster, given just four test items, selected at random from the pool, given the response pattern (#63 wrong, #23 right, #1 wrong, #38 wrong), and given the Bayesian reasoning methods we have been employing.

The decision as to when to terminate the test depends on how willing we are to make false mastery and false nonmastery decisions (type I and II errors). A type I error, α, is the probability of choosing mastery when the nonmastery alternative is really true. A type II error, β, is the probability of choosing the nonmastery alternative when the mastery alternative is really true. Most expert systems do not contain statistically based stopping rules. However, we can adopt the rules developed by Wald (1947) for the Sequential Probability Ratio Test (SPRT).

Stopping rule 1. If the ratio of the posterior probabilities of the two alternatives (mastery vs. nonmastery) derived from Bayes's theorem is greater than or equal to $(1 - \beta) \div \alpha$, then stop making observations and choose the first alternative (mastery in this context).

Stopping rule 2. If the ratio of the posterior probabilities of the two alternatives (mastery vs. nonmastery) derived from Bayes's theorem is less than or equal to $\beta \div (1 - \alpha)$, then make no more observations and choose the second alternative (nonmastery).

Continuation rule. If the ratio of the posterior probabilities of the two alternatives is neither greater than or equal to $(1 - \beta) \div \alpha$, nor less than or equal to $\beta \div (1 - \alpha)$, then take a new observation, update the posterior probabilities using Bayes's theorem, and apply the three rules once again.

Suppose that we set $\alpha = \beta = 0.05$. The threshold for the first rule is $(1 - 0.05) \div 0.05 = 0.95 \div 0.05 = 19$. The threshold for the second rule is $0.05 \div (1 - 0.05) = 0.053$. During these observations, the first three result in posterior probability ratios that fall between the two thresholds. The ratio of the posterior probabilities after the fourth observation, however, is $0.024 \div 0.976 = 0.025$, which is less than 0.053, the threshold for stopping rule 2. Therefore, we would conclude that the student is a nonmaster, knowing that we would tend to be wrong about 5% of the time, because we set β a priori at 0.05.

In summary, this example illustrates how data-based decision making can be made by a computer-based testing system, using expert systems reasoning—in particular, Bayesian reasoning—and rule quadruplets that were constructed from APTs of data derived from testing a representative sample of students who are masters and nonmasters. In effect, this approach combines Bayesian reasoning, with APT-derived probabilities of patterns for the rule base, and SPRT stopping rules. I (Frick, (1989) suggested this approach as a response to the criticism of the SPRT, in which all items are treated as if they were equally difficult. Although I demonstrated that the SPRT has high predictive validity—if used conservatively (small α's and β's)—and that test lengths can be kept fairly short (about 20 items on the average), the SPRT still requires choosing both a mastery and nonmastery level a priori. The wider the gap between the two levels, the shorter tests tend to be, and tests that are too short might be expected to result in more decision errors.

Empirical Validation of the EXSPRT

The new approach, which combines both expert systems and SPRT principles, is called EXSPRT. To investigate initially the predictive validity of this approach, extant computer-based test data were reanalyzed using APT queries such as, If student achievement status is mastery and test item pool is DAL test and item number is 25, then student answer is correct? Two test item pools were available: (a) DALTEST, a 97-item test on the structure and syntax of the Digital Authoring Language (53 administrations), and (b) COMTEST, an 85-item test on basic computer literacy and how computers functionally work (104 administrations).

Each set of test results was originally stored in a data base on an item by item basis, in the randomly selected order they were administered to an examinee. It was therefore possible to retroactively apply the EXSPRT with APT-derived rule bases generated from 50 randomly selected ad-

ministrations of each test item pool, and compare the EXSPRT decision outcomes with those reached by administration of the entire tests. Reverse APT queries were made, because we wanted to predict backward in time. In essence, the APT score is turned upside down to make a reverse query. Results of such queries about temporal patterns are determined by searching the data starting at the end and working toward the beginning of the APT score. Results were as follows:

1. If total test decision is mastery
   and test item pool is DAL test,
   25 out of 53
   then EXSPRT decision is mastery?
   24 out of 25, likelihood = 0.960.
2. If total test decision is nonmastery
   and test item pool is DAL test,
   28 out of 53
   then EXSPRT decision is nonmastery?
   26 out of 28, likelihood = 0.929.

If we combine the results from queries 1 and 2, the EXSPRT correctly predicted mastery and nonmastery decisions, based on the entire 97-item test, on 50 out of 53 administrations. The overall prediction error rate was 0.057, slightly above the expected a priori rate of 0.05 ($\alpha + \beta$). Not only did the EXSPRT predict fairly accurately, it did so with an average of 8.4 randomly selected items for mastery decisions and 9.4 items for nonmastery decisions. The overall average test length was 8.9, and EXSPRT decisions agreed with total test decisions 94.3% of the time.

Similar reverse queries were made for the computer literacy test administrations:

3. If total test decision is mastery
   and test item pool is computer literacy test,
   76 out of 104
   then EXSPRT decision is mastery?
   75 out of 76, likelihood = 0.987.
4. If total test decision is nonmastery
   and test item pool is computer literacy test,
   28 out of 104
   then EXSPRT decision is nonmastery?
   27 out of 28, likelihood = 0.964.

If these two query results are combined, the EXSPRT predicted mastery and nonmastery decisions accurately in 102 out of 104 cases, for an error rate of 0.019, which is less than the expected rate of 0.05. An average of 13.6 randomly selected items were required for EXSPRT mastery decisions and 12.3 items for nonmastery decisions. The overall average EXSPRT test length was 13.2, and EXSPRT decisions agreed with total test decisions 98.1% of the time.

Only 50 prior test administrations were selected at random for each APT-derived rule base. Larger sample sizes did not appreciably improve predictive validity for the computer literacy test. Although further validation studies are planned, it would initially appear that the EXSPRT is a strong alternative to IRT-based approaches to adaptive mastery testing, because the EXSPRT has high predictive validity and does not require such large samples of examinees in advance. Moreover, the EXSPRT reached its decisions with an average of 9 to 13 randomly selected items on the two tests studied thus far.

The EXSPRT has been extended to include an intelligent item selection procedure during a computer-based test, referred to as EXSPRT-I. Instead of selecting items randomly during a test, items are chosen on the basis of their ability to discriminate between masters and nonmasters and their compatibility with an examinee's estimated achievement level. With the EXSPRT-I, adaptive tests are even shorter, requiring an overall average of six and eight items on the two tests just described. See Frick et al. (1989) for further details, including a comparison of the EXSPRT with the one-parameter IRT model.

Although the EXSPRT has evolved considerably beyond initial analysis of patterns in time, it should be noted that it was an APT view that led to the formation of a rule base, which in turn was used by an adaptive testing system that also incorporated other features such as Bayesian inference, SPRT stopping rules and intelligent item selection procedures.

Discussion

In the broadest sense, temporal pattern recognition is not a new idea. The ability to recognize temporal patterns helps humans predict future happenings. For example, if one releases a pencil being held above the table, we can predict that under ordinary circumstances the pencil will fall.

I have added rigor to the process of temporal pattern recognition by using concepts from information theory, set theory, and probability theory, by developing a systematic method for data collection (APT scores), and by formalizing methods for aggregating results (APT queries)—all of which have been referred to here as *analysis of patterns in time* (APT). The discussion that follows focuses on other extant methodologies that are similar to but different from APT.

The Linear Models Approach

As indicated in the introduction, the linear models approach (LMA) historically has tended to dominate educational research methodology, although naturalistic or more qualitative approaches have gained status in the past decade or so. The LMA has also tended to dominate the social sciences in general. (See Frick, 1983, for further discussion of this phenomenon.) The world view in the LMA is that we measure variables

separately and then attempt to characterize their relationship with an appropriate mathematical model, where in general variable Y is some function of X. A mathematical equation is used to express the relation. In its most basic form, the equation represents a straight line in a two-dimensional Cartesian coordinate system, and the principle can be extended to many variables (including those measured at a nominal level by use of dummy coding schemes). In essence, the relation is modeled by a line surface, whether straight or curved, in n-dimensional space. When such linear relations exist among variables, then a mathematical equation with estimates of parameters (e.g., regression coefficients) is a very elegant and parsimonious way to express the relation.

In APT, the view of a relation is quite different. First, a relation occurs in time. A relation is viewed as a set of temporal patterns, not as a line surface in n-dimensional space. There is no imposition of any mathematical model in APT with respect to identifying a temporal relation. A relation is measured in APT by simply counting occurrences of relevant temporal patterns and aggregating the durations of the patterns. This may seem rather simplistic to those accustomed to the LMA, but Kendall (1973) notes,

> Before proceeding to the more advanced methods, however, we may recall that in some cases forecasting can be successfully carried out merely by watching the phenomena of interest approach. . . . Nor should we despise these simple-minded methods in the behavioural sciences. . .(p. 116)

In APT a *variable* is usually a temporal pattern. In the study of academic learning time discussed earlier, one of the variables was student engagement when direct instruction was occurring. Measures of that temporal pattern were obtained on a sample of 25 students, and a mean and standard deviation were formed in a normal manner. See column 8 of Table 1. The average probability of student engagement during direct instruction was 0.967 with a standard deviation of 0.029.

On the other hand, if an LMA is adopted, the data in columns 2 and 3 in Table 1 are correlated. The linear correlation between student engagement (EN) and direct instruction (D1) was 0.57, and the regression equation is

$$\text{EN} = 0.57 + 0.40\ (\text{DI}). \tag{1}$$

R^2 was about 0.32, and the standard interpretation is that about one third of the variance in the amount of student engagement is predictable from knowledge of the amount of direct instruction occurring. If direct instruction occurs 50% of the time, for example, then student engagement would be predicted to be about 77%. One might argue that Equation 1 contains more information than the APT results. For example, if DI = 1 (i.e., direct

instruction is occurring all the time), then EN would be predicted to be 0.97, and if DI occurs none of the time (DI = 0), then EN would be expected to be 0.57. These findings are essentially the same as those from APT. It turns out that this is a fortuitous coincidence because the joint occurrence of direct instruction and student nonengagment was nearly 0. Therefore, the joint occurrence of direct instruction and student engagement was nearly equal to the occurrence of direct instruction. If a relation is deterministic, then APT and LMA results will be consistent.

If a relation is stochastic, however, results from the two approaches will differ. To illustrate the differences more clearly, let us try to predict direct instruction given knowledge of student engagement. The APT probability of DI was estimated to be 0.561 if students were engaged, and 0.058 if students were not engaged (not reported in Table 1 here, but derivable—see Frick, 1983). The linear regression equation fitted to the sample data, however, is

$$DI = -.176 + .819\ (EN). \tag{2}$$

Making a similar interpretation, if student engagement is 100%, then the predicted amount of direct instruction is 0.643. If student engagement is not occurring, the estimate of the proportion of direct instruction is −0.176. Clearly, the APT and LMA results differ here. Moreover, in the LMA the proportion of DI is predicted to be negative when EN is less than 21.5%, which makes no sense. The reason for this is that beta weights and constants in regression analysis are not constrained to lie between 0 and 1 inclusively as are probabilities or proportions. One could argue that we have simply extrapolated too far in the LMA, but that is not the essence of the problem. Rather, the discrepancies are due to different assumptions about the nature of a relation.

A relation is deterministic if each category in the first classification is associated with only one category in the second classification. A relation is stochastic if a category from the first classification is associated with two or more categories from the second. That is, a category in the second classification is not uniquely associated with a category in the first classification in a stochastic relation. (See Frick; 1983, p. 10, for further details.)

In general, it is not possible to predict joint probabilities (and hence determine conditional probabilities) solely from knowledge of distributions of marginal probabilities, except in a few special cases where certain joint probabilities are equal to their respective marginals (Frick, 1983). Furthermore, when considering the prediction of temporal sequences from aggregate information about marginals only, it simply cannot be done (cf. Blossfeld, Hamerle, & Mayer, 1989; Tuma & Hannan, 1984). In other words, one cannot go from a grosser level of measurement to a finer level. If data are collected with APT in mind, it is always possible to treat those

data with the LMA if desired. The converse does not obtain, except when relations are deterministic.

Time series analysis. Time series analysis (TSA) is one variation of the LMA. A time series is "a series of observations, x_j (n); $j = 1, \ldots, p$; $n = 1, \ldots N$, made sequentially through time. Here j indexes the different measurements made at each time point n" (Hannan, 1970, p. 3). In terms of the basic idea, TSA and APT are similar indeed. However, the fundamental difference in the conception of relation still obtains. In TSA the aim is to estimate parameters of a mathematical model (i.e., equation or set of equations) that provide a good fit to the temporal data. In APT temporal patterns are counted and their durations aggregated, with no imposition of a mathematical model to describe the relation. In APT the temporal pattern is the model.

Other kinds of analytical methods in the LMA. Other LMA methods that attempt to estimate parameters of mathematical equations to explain covariation of multiple measures include path analysis, canonical analysis, factor analysis, discriminant analysis, and of course analysis of variance, covariance, and multivariate analysis of variance. APT differs from all of these in that relations are not assumed to be deterministic. In APT, a relation is not constrained to be a function—in the set-theoretic sense of function—as is the case in the LMA and all its analytic methods that assume that relations are functions (cf. Coombs et al., 1970, pp. 351 – 371).

Event History Analysis

As pointed out by a statistician who was asked by the editor of the AERJ to review an earlier version of this article, APT is similar in conception to numerous methodologies referred to collectively as "event history analysis" (cf. Blossfeld et al., 1989; Tuma & Hannan, 1984).

> By "event history analysis" we mean statistical methods used to analyze time intervals between successive state transitions or events. The number of states occupied by the analyzed units are finite, but the events may occur at any point in time. Consequently, in event history analyses statistical methods for analyzing stochastic processes with discrete states and continuous time are used. (Blossfeld et al., 1989, p. 11)

The notion of event history analysis is not a new idea. For example, Coleman (1964) discussed the use of Markov models to study social processes. Coombs et al. (1970) also suggested Markov models and di-graphs as means of studying change processes in the field of psychology. As educational examples, Bellack, Kliebard, Hyman, and Smith (1966), Flanders (1970), and Collett and Semmel (1973) have conducted studies where sequential student/teacher behavior patterns were of interest.

The Markov model. If only one classification is of interest in APT,

then APT bears some resemblance to the Markov model. In such a model, a state-space approach is postulated, where the process studied is characterized as being in only one of a number of possible states at a given moment in time. A set of states is considered, $S = \{S_1, \ldots, S_m\}$, at each moment in time, $\{T_1, \ldots, T_n\}$. Of particular interest are the transitional probabilities between successive states. In the Markov model it is assumed that the transitional probabilities between successive states are unaffected by the history of the stochastic process. In APT this assumption is made also. However, in APT temporal patterns longer than two-stage sequences can be investigated. Moreover, in APT multiple classifications can be considered simultaneously not just one as in the Markov model. (See Frick, 1983, for further discussion, including comparison of APT to multivariate contingency analysis—Goodman, 1978.)

Other models. Blossfeld et al. (1989) and Tuma and Hannan (1984) list numerous models for the analysis of event history data, including survivor functions, hazard functions, competing risk models, log-logistic models, and so on. As with the LMA, in each case some mathematical model is postulated in an attempt to describe a stochastic process (with the exception of the Life Table Method). I do not question the value of these various approaches. For example, it certainly is useful to be able to estimate the probability that a patient will live at least 5 years after receiving some treatment for cancer. Indeed, different treatments can be evaluated by comparing their respective survivor functions.

It is clear that APT scores are amendable to such analytical methods if desired. It should be noted also that a typical goal of event history analysis is to estimate the probability of how long a given state will last before it changes into a different state. For example, in an event history analysis of the ALT data one might ask questions such as, If direct instruction is occurring, what is the probability that a student will remain engaged for at least 5 minutes without going off task? In APT the question would be phrased, At any point in time that we observe academic activities in education, and direct instruction is occurring at that time, what is the likelihood that a student is engaged in that activity?

What is significant, however, in nearly all of the methods discussed in these two excellent volumes (Blossfeld et al., 1989; Tuma & Hannan, 1984), is that some kind of mathematical model is being assumed. For example, Blossfeld et al. (1989) assert, "After the construction of a statistical model for the event history under discussion, the unknown parameters have to be estimated from the data" (p. 64).

In APT this is not the case. In APT there is no concern for discovering some equation (mathematical model) that will predict a regularity in a stochastic process. Rather, in APT the onus is on the investigator to search for patterns by examining APT scores and forming queries. Presumably the search is guided by some hypotheses or questions. The results of these

queries can be tabulated, as was done in Tables 1 and 2. In APT we simply ask, What is the probability that a particular kind of event will occur, given the specification of a pattern of antecedent events? Each query is a model, so to speak, and the model is not a mathematical equation but rather the specification of a temporal pattern.

Causality

Causal conclusions are not warranted from results of APT queries alone. For example, it is tempting to conclude that direct instruction causes student engagement with academic tasks. Clearly direct instruction is not the cause, as student engagement also occurs in the absence of direct instruction. Direct instruction may be a factor that is associated with increased student engagement, or it could be a proxy for something else not classified by the current observation scheme.

Perhaps a clearer example is the temporal pattern of dawn-then-sunrise. We would not conclude that dawn causes sunrise, even though the pattern is highly predictable. We know of course, at least since Copernicus, that this temporal pattern is due to the earth's spin about its axis as it revolves around our sun.

Thus, co-occurrence or sequential occurrence of events do not imply that one event causes another. Normally,the best way to determine causation is to conduct an experiment, if possible, and manipulate one factor and observe its effect on the other while trying to randomize or control for the effects of any other factors.

For example, in a follow-up study to the one described earlier, teachers were asked to try to increase direct instruction during the spring semester, after observing them and their students during a 2-month baseline during the fall semester. Not every teacher did so. In fact, a number of teachers decreased the amount of direct instruction in their classrooms in the spring semester. In the 11 cases where teachers decreased the proportion of direct instruction, the proportion of student engagement also decreased in 8 of those cases. In 11 of 14 cases where teachers actually did increase the proportion of direct instruction in their classes, the proportion of student engagement also increased. In those few cases where the trend did not occur, student engagement tended to be very high in the fall semester, and it is possible that ceiling effects prevented a further increase. Given the results of these 25 case studies, it would appear that direct instruction is often one causal factor in determining elementary student engagement in academic tasks. (See Rieth & Frick, 1983, for further details.)

Generalizability of APT Query Results

As with any descriptive measures, the generalizability of APT query results depends on the relative number of systems observed, how they are selected, when they are observed, and to what population of systems

generalizations are to be made. The same issues of sampling strategies (both within sampling units across time as well as of sampling units), observer accuracy in coding, and so forth, that arise in normal survey or observational research still apply to the design of research studies that employ APT methodology. The reader is warned that collecting event history data is often time consuming and expensive (cf. Blossfeld, et al., 1989).

One can estimate mean probabilities of occurrences of temporal paths, as was done in the academic learning time study; and confidence intervals can be estimated by ordinary statistical methods. These probabilities can be based on relative frequencies of event changes, or they can be based on relative duration of temporal patterns, depending on which is more appropriate for the questions being asked. (See Frick, 1983, 1988, for further details.)

Conclusion

Analysis of patterns in time, as a general notion, is not a new idea. The fruitfulness of this approach was illustrated by applying APT methodology to a classroom observational study. It was also exemplified by use of APT methodology to develop a rule base about test items that was in turn referenced by a computer-based adaptive testing system in making decisions about student mastery of an educational objective.

APT was compared with numerous extant methodologies, including the linear models approach and event history analysis. The fundamental difference between APT and these other approaches is that no particular mathematical model is assumed in APT. In APT a model is viewed simply as a temporal pattern, whereas in most other approaches parameters of a mathematical model are estimated from data in which variables are measured separately. Moreover, in APT probabilities of temporal patterns are estimated by relative frequency and duration.

Many statisticians may view APT as a simple and elementary technique of temporal data description. Nonetheless, the two research studies summarized earlier do indicate the power of a simple but useful idea.

As a further example, APT queries and their results may be used to form rules for expert systems that become part of an intelligent computer-based instructional system. Such a system theoretically can optimize student learning by recommending instructional sequences (i.e., temporal patterns) that have high probabilities of resulting in student mastery. In other words, APT-based decision making by a computer program can provide an empirical foundation for artificial intelligence.

As a final note, APT is another tool for doing research. Clearly, APT is not suitable for all problems. Researchers should choose methods that fit the problems addressed, not vice versa. Perhaps, though, APT will encourage educational researchers to view some existing problems with a different mind set. If that happens, then this article has achieved its goal.

Notes

[1]The APT methodology was previously termed *non-metric temporal path analysis* (NTPA). Although the previous terminology was logical given the conception of the methodology, it tended to be confused with traditional path analysis, with which it bears no relation. Thus, the name has been changed to *analysis of patterns in time* (APT). The methodology has not been changed, only the name.

References

Bellack A.A., Kliebard, H.M., Hyman, R.T., & Smith, F.L. (1966). *The language of the classroom.* New York: Teachers College Press.

Blossfeld, H.-P., Hamerle, A., & Mayer, K.U. (1989). *Event history analysis: Statistical theory and application in the social sciences.* Hillsdale, NJ: Erlbaum.

Coleman, J.S. (1964). *Introduction to mathematical sociology.* New York: Free Press.

Collett, L.S., & Semmel, M.I. (1973). *The analysis of sequential behavior in classrooms and social environments: Problems and proposed solutions.* Bloomington, IN: Indiana University, Center for Innovation in Teaching the Handicapped.

Coombs, C.H., Dawes, R.M., & Tversky, A. (1970). *Mathematical psychology.* Englewood Cliffs, NJ: Prentice Hall.

Flanders, N.A. (1970). *Analyzing teacher behavior.* Reading, MA: Addison-Wesley.

Frick, T.W. (1983). *Non-metric temporal path analysis: An alternative to the linear models approach for verification of stochastic educational relations.* Unpublished doctoral dissertation, Indiana University School of Education.

Frick, T.W. (1988, August). *Non-metric temporal path analysis: A method by which tutorial systems can learn through inquiry.* Paper presented at the Fourth International Conference on System Research, Cybernetics and Informatics, Baden-Baden, West Germany.

Frick, T.W. (1989). Bayesian adaptation during computer-based testing and computer-guided practice exercises. *Journal of Educational Computing Research, 5*(1). 89–114.

Frick, T.W., Plew, G.T., & Luk, H.-K. (1989, March). EXSPRT: *An expert systems approach to computer-based adaptive testing.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Frick, T.W. & Rieth, H.J. (1981). *ALTOS observer reference manual: Academic Learning Time Observation System.* Bloomington, IN: Indiana University, Center for Innovation in Teaching the Handicapped.

Goodman, L.A. (1978). *Analyzing qualitative/categorical data.* Cambridge, MA: Abt Books.

Guba, E.G., & Lincoln, Y.S. (1981). *Effective Evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Hannan, E.J. (1970). *Multiple time series.* New York: Wiley.

Kendall, G.W. (1973). *Time-series.* New York: Hafner Press.

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

Maccia, E.S., & Maccia, G.S. (1976). *On the contribution of general systems theory to educational research.* Paper presented at the Society for General Systems Research, Boston.

Maccia, G.S., & Maccia, E.S. (1966). *Development of educational theory derived from three educational theory models.* (Final Report, Project No. 5-0638). Washington, DC: U.S. Office of Education.

Mager, R. (1973). *Measuring instructional intent.* Belmont, CA: Fearon-Pittman.

Rieth, H.J., & Frick, T.W. (1982). *An analysis of academic learning time (ALT) of mildly handicapped students in special education service delivery systems: Initial report on classroom process variables.* Bloomington, IN: Indiana University, Center for Innovation in Teaching the Handicapped.

Rieth, H.J., & Frick, T.W. (1983). *An analysis of the impact of different service delivery systems on the achievement of mildly handicapped children* (Final Report, Project No. 443CH00168). Washington, DC: U.S. Department of Education.

Schmitt, S. (1969). *Measuring uncertainty.* Reading, MA: Addison-Wesley.

Tuma, N.B., & Hannan, M.T. (1984). *Social dynamics: Models and methods.* Orlando, FL: Academic Press.

Wald, A. (1947). *Sequential analysis.* New York: Wiley.

Weiss, D., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21,* 361–375.