DOCUMENT RESUME

ED 092 119 IR 000 678

AUTHOR Dyrenfurth, Michael; And Others
TITLE The Use of Physiological Indices in Simulation Research: A Report on Project CORES (Covert and Overt Responses to Educational Simulations). A Symposium.
INSTITUTION Bowling Green State Univ., Ohio.
PUB DATE Apr 74
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974); For related document see IR 000 677

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Administrative Personnel; *Administrator Education; Concept Formation; Measurement Instruments; *Physiology; Research Methodology; Simulated Environment; *Simulation
IDENTIFIERS *Project CORES

ABSTRACT

In two separate reports the founding and set up of Project CORES was outlined, and then a specific research project was described. Project CORES began in the efforts of three men who felt a more systematic investigation of simulation effects was needed. The criteria felt most sensitive were the physiological activities of galvanic skin potential and heart rate. A simulation chamber was constructed in the form of a small office with a screen where a program could be projected. The research project utilized the Project CORES facility with an added measure of concept meanings, the semantic differential technique. The concept meanings formed a pre- and posttest for a single group, N=32, going through the simulation sessions. A positive relationship between the degree of involvement as indicated by physiological activity and change in concept meaning was evidenced. (WH)

6.20
P28

The use of physiological indices in simulation research: A report on Project CORES (covert and overt responses to educational simulations)

A Symposium

Michael Dyrenfurth, Chairman
Valley City State College, ND

Participants

Richard Coulter
North Union Local Schools, Ohio

Michael Dyrenfurth
Valley City State College

Gene Poor
Bowling Green State University

Bill Reynolds
Bowling Green State University

Discussant

Jack Culbertson
University Council for Educational Administration, Columbus, Ohio

Presented to American Educational Research Association Annual Meeting, Chicago, 1974

PROJECT CORES: AN OVERVIEW

This symposium introduces the notion of the use of physiological responses, such as galvanic skin responses, heart rate, and respiration rate, as indices of involvement in simulation. As such the technique provides a useful tool for the developer of simulations to assess the impact of his creations. Uses in both the formative and summative domains of evaluation are possible.

Furthermore, the project postulates a comprehensive systems model of simulation and its components. Through dialog and further research the relationship among the components will be clarified, thereby aiding the development and application of new simulations.

HISTORY & BACKGROUND OF PROJECT CORES

The project started as three individual dissertation efforts for Gene Poor, Richard Coulter, and myself. As a result of what I assume to be a typical guidance and counseling process, two things became clear:

A. That the field of educational administration suffered from a surfeit of surveys and role studies.

B. That there was a need for more knowledge centering on the use of simulation techniques in educational administration.

Once these two ideas crystalized we each launched into our initial review of the literature. The original review of literature in the field of simulation indicated very little experimentally derived and substantiated evidence as to the effect of simulations. Hunches and intuitive interpretations were rampant however. Additional research also pointed out that the mechanisms and variables of simulation systems were equally unsupported and clarified by empirical research.

The next crucial step was the identification of the possibility of using physiological indices as indicators of involvement. From there each of us went our separate routes in defining our problems and research. We did, however, collectively pursue some funding from the graduate school at Bowling Green State University. The resultant funding in conjunction with our procurement capabilities allowed us to procede in the manner to be detailed.

A coordinated project was initiated by Bowling Green State University's Division of Educational Administration. The project's ongoing purposes are to define important variables in the simulation system and to develop implementations thereof. Three initial investigations comprised the first phase of the project:

I. Dr. Gene Poor's investigation of the simulation vehicle, ie., the media used to convey the simulation program content to to the simulation participant (1973).

II. Dr. Richard Coulter's investigation of the relationships between simulation responses, background, and experiential variables (1973).

III. Dr. Michael Dyrenfurth's investigation of the relationships between the simulation participant's physiological reactions and the effect of the simulation (1973).

IV. Several additional studies presently being pursued will be described by Dr. Bill Reynolds.

THE STUDIES' COMMON CORE

Because we three were the first at BGSU to pursue this particular thrust, a significant portion of our effort was devoted to establish the hardware necessary for the research effort. Four major areas of commonality need to be described. The simulator, the program, the sample, and the physiological measures.

Simulator

The dynamic administration simulator (DAS) depicted in figure 1 is a cubicle approximately seven foot square. Built to resemble a typical office, the DAS was also designed as a tool to achieve environmental control of sound and temperature. The walls are attractively panelled and treated with internal insulation. The suspended ceiling consists of insulating tile and contains lighting and ventilation provisions. The floor is carpeted.

The contents of the DAS include; a desk, a desk chair, and a guest chair. One wall, directly in front of the desk, is fitted with a large rear projection screen that is flanked by a one-way mirror and a set of stereo speakers.

The equipment used for injection of the program into the DAS includes a 16mm film projector, two 35mm slide projectors, a telephone, an audio system, and light level controls. The output and monitoring equipment used in the first generation DAS included a telephone, a keyboard response panel, a one-way mirror and physiological monitoring apparatus.

Program

The term, program, refers to the sequence and set of inputs describing the simulated situation. These inputs were visually pro-

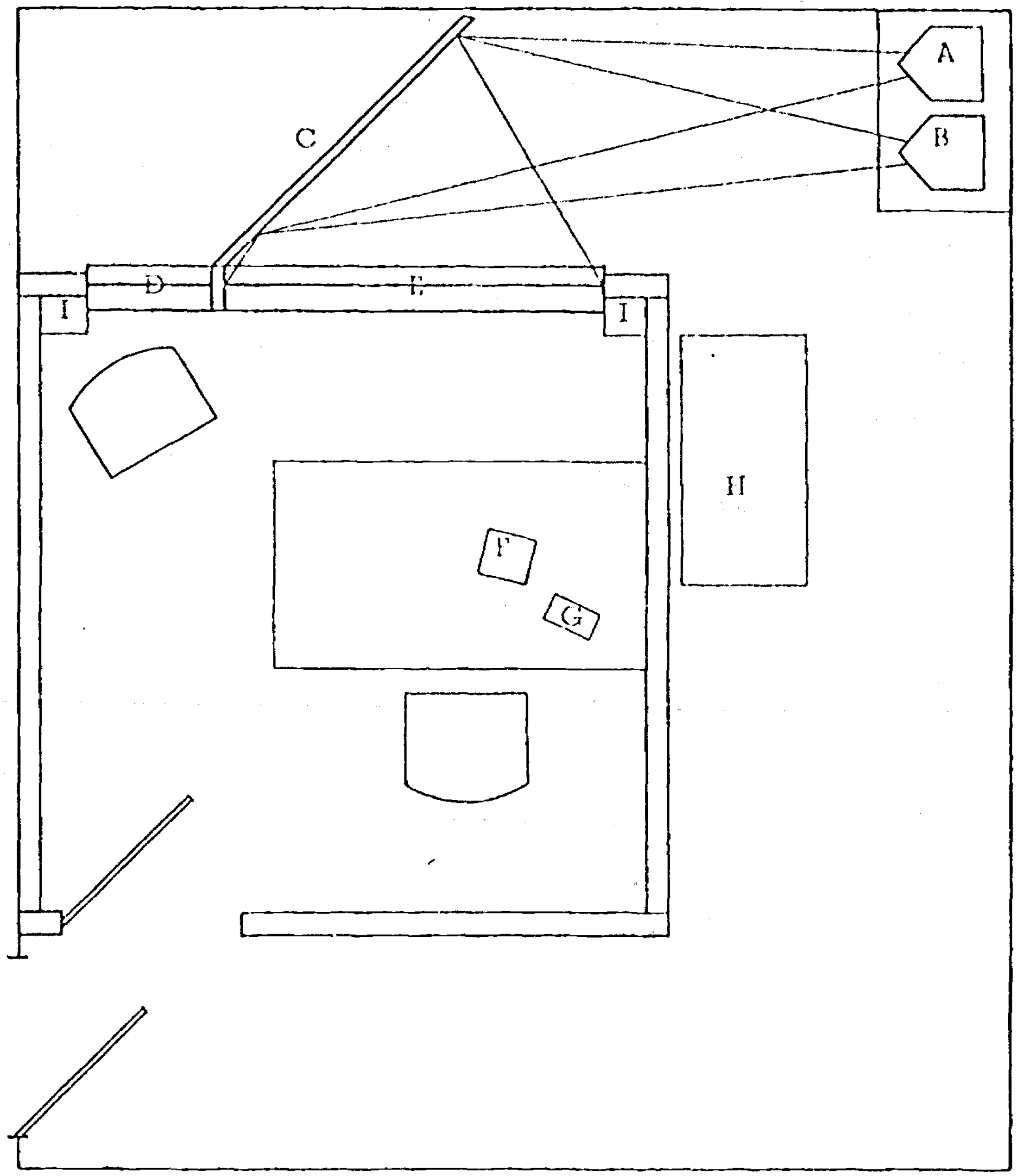

Figure 1. Dynamic Administration Simulator and Facility.

A - Slide projector
B - Film projector
C - Rear projection mirror
D - One-way mirror
E - Rear projection screen
F - Telephone
G - Response panel
H - Simulator controls and polygraph
I - Speakers

jected using either 16mm color film or 35mm color slides. Audio was provided by both a magnetic soundtrack on the film and by the telephone on the subject's desk. At key times during this program pauses were introduced to allow the participant to respond to the problem incident as he chooses. The responses consisted of a choice selected from given ones projected on the screen. The participant indicated his choice by pressing the appropriate button on the response panel on his desk.

The program used in this study was based on and developed from selected incidents produced by Dr. B. J. Reynolds. The original set of incidents, that is, occurances within the principal's day, were part of the Alta Vista Package. As conceived by Dr. Reynolds, Alta Vista is intended to be a "total simulation", that is, a simulation with a large scale, wide angle, multiple perspective of a school system's operation. A script dialog was then developed to convert the incidents into action. Subsequently, using the help of an amateur theatrical group, the script was then filmed to yield the visuals required for projection.

The program's story line involves the experiences of a secondary school principal as he handles a morning of events, all intimately involved with a teacher walkout. To compound the situation, the principal is expected to carry out the superintendent's policy of keeping the school open. Further difficulties include an injured student, conniving teachers, and union pressures. After each participant is steeped in the situation he is presented with a set of five alternatives and is suddenly faced with the necessity of selecting an appropriate response.

The Population and Sample

The population was defined to include those educational administrators in the northwest section of Ohio that cluster around Bowling Green State University. Educational administrator was defined to mean all those presently involved in, or preparing for, administration or administration and teaching in public primary, secondary, and post-secondary institutions. It was assumed that students registered in the administrative courses offered by Bowling Green State University were representative of this population.

Because of the nature of physiological research and because of the general unfamiliarity of educators with the techniques employed, it appeared difficult to resort to the more common and respected random sampling techniques. Accordingly, the sampling method used was to call for volunteers from the above mentioned classes. It was attempted to obtain a total sample size of 35. The actual useable sample size was 32.

Furthermore, the difficulty or perhaps even the impossibility of establishing a control group, without any treatment, precluded the use of a control group design. It was recognized that the chosen design limited the extent of inferences based upon the study's findings. However, because of the exploratory nature of the investigation as well as the need for a study more inclusive than a strictly experimental investigation of just one variable, the restrictions on inference were considered less important than the advantages gained by the broad look at activity.

Inprocess Physiological Instrumentation

Overview

The inprocess measures consisted of continuous monitoring of two basic indicators of physiological activity: galvanic skin potential (GSP) and heart rate (HR). Each measure was obtained through the use of electrodes sensing the electric signals associated with the physiological phenomena concerned. Non-clinical electrode positioning was employed. Skin preparation followed the recommended procedure, namely slight skin abrasion, decornification, for the GSP reference electrode and alcohol cleansing for the others.

Selection of Measures

It was necessary to select physiological indicators that would provide clear and relatively unambiguous information as to a subject's covert reactions to simulation. Furthermore, it was important that these measures be available continuously during the treatment.

The major physiological variables associated with the arousal and emotions generated by stressful situations can be categorized as pertaining to: (a) Electrical properties of the skin, (b) Characteristics of the cardiovascular system, and (c) Respiratory phenomena (Grossman, 1967).

Because of the exploratory nature of this study, it was decided to focus on the three major categories of variables listed by Grossman. A further review of the physiological and psychophysiological literature was conducted to determine which measure to use in each category. This review indicated that in order to obtain a reliable measure more than one indicator had to be used (Sternbach, 1966).

The researcher therefore set out to use three measures, one pertaining to each of Grossman's (1967) major categories. However, the review indicated (Schnore, 1959), that not all physiological measures would be equally useful in differentiating between participant activity levels. Perhaps the lesser usefulness, for this study at least, of respiratory measures in differentiating emotional level, is due to the fact that they are partially subject to voluntary control (Grossman, 1967). Because of these findings it was decided to forego the use of any respiratory measure.

Fortunately, convenient measures in each of the remaining two major categories meet the criterion. Grossman (1967) claims that "visceral responses" which include HR, GSR, and GSP are not subject to voluntary regulation. Thus it remained only to select the actual measures to be used as indicators of the activity of: (a) The electrical phenomena of the skin, and (b) The cardiovascular system. With respect to the former, GSP was selected despite the fact that GSR is the most widely used index (Lader et. al., 1966). The initial reason for choosing GSP was that GSR could not be measured with the apparatus available for this investigation. This academically unsupportable reason was fortunately augmented by the findings of Lader et. al., (1966), who reported that the "psycho-galvic response"; their generic name for electrical skin phenomena, can be obtained using either both GSR or GSP methods.

Of the several measures available to indicate cardiovascular activity, HR was selected for use in this study. While, as indicated previously, multiple physiological measures were necessary to obtain a reliable indication of activation, caution had to be exercised since two highly correlated measures would not augment reliability. HR is reported as being independent of Palmar conductance. Palmar conductance is the reciprocal of GSR, a measure highly correlated with GSP, (Lader et. al., 1966, citing Lacey and Lacey, 1958). Furthermore, just as Shapiro et. al., (1964) reported on the usefulness of the GSP as an index of the degree of activation, Schnore (1959) clearly includes HR with the variables he labels as consistent differentiators between high and low arousal levels.

Details of the Technique

While some of the specific details of the instrumentation precedures were able to be determined by pilot runs, many other details were abstracted from the literature. HR, for example, presented little problem in electrode placement. The signals associated with this activity are so strong that they are easily obtainable from many parts of the body. Electro-dermal phenomena, however, are not as easily measured. Much argument exists over proper techniques (Lader, et. al., 1966). Venables and Sayer (1963) were of major assistance in determining the specific location, of palm and arm of GSP electrodes. They also recommend skin abrasion for the inactive site.

Apparatus

The inprocess measures were recorded using a Beckman model RB dynograph with four available channels and a time event marker. Signals were sensed by Beckman miniature skin electrodes using a commercial paste. Both the participant's physiological reactions and his choice of response to the problem by means of a home built indicator box were recorded by this apparatus.

The physical design of the apparatus involved as inconspicuous connections and wiring as possible. The connections and wiring were arranged so as to allow the participant some freedom of movement within his chair. Other than the electrodes and wires, all actual instrumentation was mounted outside the DAS and was not seen by the participant until after the experiment. All experimentation took place in a shielded room with considerable accoustical control. Environmental variables (temperature and humidity) were controlled by the plant's air-conditioning system.

Reliability of the Inprocess Measures

With the basic physiological procedures selected, two sets of pilot runs were conducted to determine such things as repeatability, rate of adaption, (the tendency for a stimulus to produce less of a response near the end of an experimental period than at the beginning of the period) and the exact technique and locations to employ.

Control of motion artifact was crucial to the interpretation of both HR and GSP. After some trial and error cycles, a set of electrode positions, exhibiting a minimum of motion artifact, was discovered. The preferred positions for the two GSP electrodes were: (a) The left hand palm on the fleshy part at the base of the thumb, near the point where the axis of the thumb and first finger intersect, and (b) on the left upper arm about two to three inches above the elbow on the inside surface. HR was sensed using three electrodes: One reference electrode on the right upper arm, about two to three inches above the elbows on the inside surface, and two active electrodes, one each left and right, high on the upper arm's outside surface, about two inches below the shoulder. It was not possible to completely control motion artifact by electrode placement and recording methods, but the remainder was controlled by means of the analytical procedures employed.

One additional fact was determined during the pilot runs. The independence of the two channels was demonstrated. There was no observable distortion in channel one (GSP) when channel two (HR) was introduced and vice versa.

The second set of pilot runs investigated the reliability of the GSP and HR measures in terms of the experimenter's consistency in electrode placement, application, instrument operation, and trace interpretation. Because the "arousal potential" of a simulation is less each time a participant repeats the exposure, the repeated measures during treatment method of determining reliability could not be used. Consequently, reliability was determined using a no-stimulus method. It was assumed that because the only difference would be in the level of response, a suitable indicator of reliability would result. Each volunteer, N = 14, was subjected to three sessions that paralleled the experimental procedures except that no form of stimulus was given. This set of pilot runs consisted of attaching GSP and HR electrodes to the volunteer and then seating him in the DAS. The volunteer remained there for ten minutes, during which time no stimuli of any sort were injected. This process was repeated for each of the three sessions.

Of the three sessions, only the second and third were scored. The first, which was considered an aclimatization session, was discarded to avoid the novelty effect artifically lowering the reliability estimates. The remaining two sessions yielded two charts or recordings, for each participant. Each chart was then scored to yield indices for HR, galvanic skin potential amplitude (GSPA), and galvanic skin potential frequency (GSPF).

Each of the indices obtained were then plotted to determine significant trends. This was done visually, primarily to inspect for evidence of adaptation. Numerical inspection for adaptation was done by using correlated tests to examine the differences between beginning and ending means. From this data, it was determined that there was no significant adaptation.

The data, was also used to correlate the participants' run two scores with their run three scores. A separate Pearson product moment correlation was calculated for each of the following indices: HR, GSPA, and GSPF. Because of the small sample size, Spearman's rank order correlation was also calculated. In addition, a scattergram was plotted for each correlation to determine linearity. This condition appeared to be satisfied. Table 1 presents the results of the reliability correlations. These concluded that the correlations were strong enough to warrant use of the developed technique.

Table 1
Reliability Correlations, Pilot Run Two Versus Run Three

Physiological Indices	Correlations[a]			
	Pearson r	t[b]	Spearman r	t[b]
GSPA	.88	5.86*	.75	3.59*
GSPF	.85	5.10*	.78	3.94*
HR	.73	3.38*	.68	2.93*

a. N=12, the responses of two participants were discarded as mavericks b. Calculated using t test by Tate (1955, p. 469)
* Significant at the .01 level, one tail test

BIBLIOGRAPHY

Coulter, Richard. An investigation of relationships between overt-covert responses to simulation age, experience, and education of educational administrators. (Doctoral dissertation, Bowling Green State University) Ann Arbor: 1973. No. 1519-a.

Dyrenfurth, Michael. Investigation of relationship between participant involvement in simulation as evidenced by physiological activity, and change in concept meaning. (Doctoral dissertation, Bowling Green State University) Ann Arbor: 1973. No. 1523-a.

Grossman, S. P. A textbook of physiological psychology. New York: Wiley, 1967.

Lacey, J. I. & Lacey, B. C. Verification and extension of the principle of autonomic response-stereotypy. American Journal of Psychology, 1958, 71, 50-73.

Lader, M. H. & Wing, L. Physiological measures, sedative drugs, and morbid anxiety. New York: Oxford University Press, 1966.

Poor, Gene. The effects of static and dynamic simulation programs on physiological activity of educational administrators. (Doctoral dissertation, Bowling Green State University) Ann Arbor: 1973. No. 538-a.

Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. Journal of Experimental Psychology, 1959, 58(2), 117-128.

Shapiro, D. & Leiderman, P. H. Studies on the galvanic skin potential level: Some statistical properties. Journal of Psychosomatic Research, 1964, 7, 269-275.

Sternbach, R. A. Principles of psycho-physiology. New York: Academic Press, 1966.

Tate, M. W. Statistics in education. New York: Macmillan, 1955.

Venables, P. H. & Sayer, E. On the measurements of the level of skin potential. British Journal of Psychology, 1963, 53, 251-260.

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN INVOLVEMENT IN SIMULATION, AS EVIDENCED BY PHYSIOLOGICAL INDICES, AND CHANGE IN CONCEPT MEANING[1]

Introduction.

It was the intent of this study to investigate simulation and attempt to identify some crucial factors, an awareness of which is necessary in order to develop and utilize the technique effectively. The purpose of the study can, therefore, be restated as seeking to contribute to the knowledge base of simulation theory and application.

When assessing the effects of simulation, researchers have usually concentrated on the quantity of learning as a criterion. Quantity, as an output measure, was frequently determined using standardized tests or researcher developed tests. In either case, because of their unidimensionality, that is, tests developed or selected to measure a particular thing, their sensitivity to fine changes in the quality, not necessarily the quantity, of knowledge or cognitive structure, is suspect. Perhaps another technique would be more sensitive? Because of its multi-dimensionality, it was postulated that concept meaning, as measured by the semantic differential technique (Osgood, Suci, & Tannenbaum, 1967), fulfilled the need for a more sensitive output measure.

Proponents of the simulation technique often imply that involvement is one, if not the, significant aspect that results in participant learning (Beck & Monroe, 1969a; Cruickshank & Broadbent, 1968; Twelker, 1969). This claim for involvement is related to such other claims as relevance, realism, activity, and reward. By virtue of its centrality to the educative applications of simulation, it was proposed that the claim of involvement was worth investigating.

---

[1]Michael Dyrenfurth, Chairman, Department of Industrial Education, Valley City State College, Valley City, ND 58072.

Involvement, if it in fact exists, by definition must occur DURING the simulation experience. Furthermore, if involvement occurs during the process of simulation then it seems more appropriate and direct to measure it during the process than after the fact with retrospective questions and self-reporting instruments. It follows then, that if this involvement is generally inherent in simulation participants, that this will be demonstrated by their physiological reactions during the simulation. The problem investigated stemmed from this premise.

The Problem

The general problem was to determine the relationships between involvement, as manifested by physiological activity during simulation, and change in concept meaning.

Research Hypothesis

Simulation participants with a high level of physiological activity during simulation will exhibit more change in concept meaning than participants with a low physiological activity level. In testable form, it can be stated that there will be a positive relationship between physiological activity during simulation and change in concept meaning.

Significance of the Study

The study's significance rests on its contribution to the knowledge of what occurs during simulation and how this relates to the technique's effects. Acceptance of the research hypothesis would provide the developer of future simulations with a criterion for determining his creation's usefulness in affecting the relevant factor structures of his participants and this may lead to a more effective way of evaluating the outcomes of simulation.

THE PROCEDURES

Overview of Procedures

This study employed an experimental group, N=32, to obtain the data necessary to test the hypothesis. The experimental group was subjected to pre and post-treatment measures of concept meaning. The treatment was the simulated situation. During the simulation the participants' physiological responses were continuously monitored and recorded. The ultimate test of the problem was pursued through correlational statistics. Participant change in concept meaning was correlated with the level of physiological activity as measured by the various physiological indices. This allowed determination of an overall index of relationship which was predicted to be significant and in the positive direction.

The Design

A modified single group quasi-experimental design was employed, namely, a one group pretest-posttest design (Campbell and Stanley, 1963). The design was modified by superimposing O_2 on X_1 that is, providing observation during treatment. This alteration was necessary because of the need to indicate concurrent treatment and monitoring.

Treatment

The simulator and its program have already been described as have the inprocess measures of physiological activity.

The participants in the experimental group underwent treatment involving the simulation of selected aspects of a secondary school principal's activity. A systems paradigm of this simulation and its context is provided by figure 1. Three inputs formed the basic resources of the system. Two of these, the program and the simulator, combined to form the model abstracted from reality.

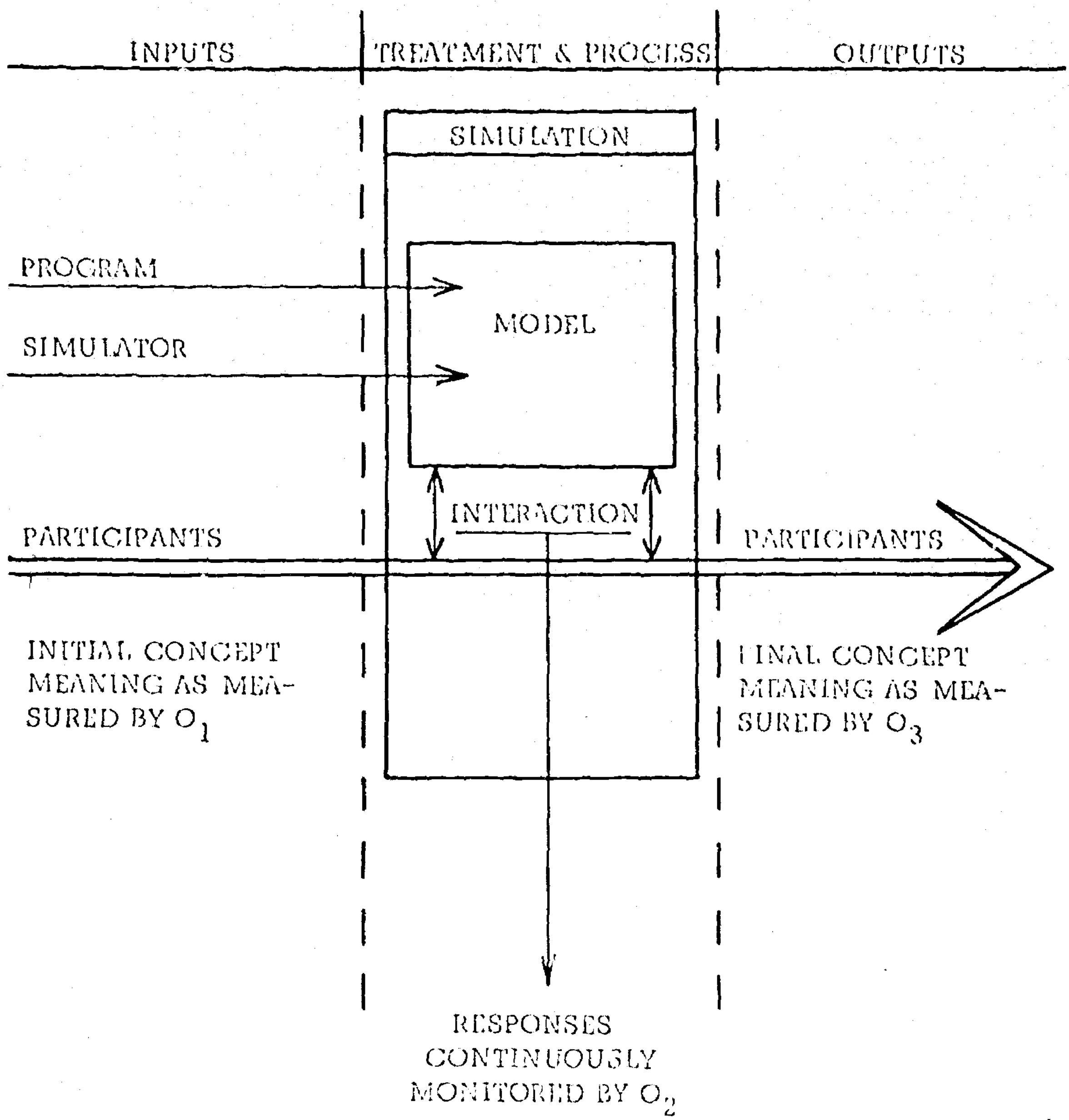

Figure 1. Systems view of simulation

The participants, as the third input, interacted individually with the model. This interaction became the actual simulation.

Output Instrumentation

A review (Dyrenfurth, 1973) indicated that a major weakness of simulation evaluations consisted of using criterion measures when some justified questions might be raised as to which criteria should be employed (Cherryholmes, 1966). The question of which criteria, if any, were to be used in this investigation was equally difficult to resolve. In fact, under the existing conditions of uncertainity as to what effect the simulation will have, the first step was to establish the fact the something" has changed as a result of the participant's exposure to simulation.

A possible second step, one based on a specific set of values, could then define the changes that occurred as "better" or "worse" than others. It was the researcher's contention, however, that the present state of knowledge pertaining to simulation did not permit the second stage to be applied. Accordingly, only the existence of change and its amount was dealt with in this study.

Change as mentioned in the preceding paragraph requires elaboration and clarification. Change of what? Attitude? Knowledge? Again, the determining factor was the formative stage the investigator found himself in. While change of attitude towards the principalship, for example, could have easily been justified on the basis of the program content, it was felt that this would repeat the error identified by the review, the error being the selection of a particular criterion while using an inadequate basis of operation. As a result of this decision a composite measure was sought, ie., one that would encompass change of many kinds. It was mandatory that this measure, for example, be able to detect both cognitive and affective change that might result from the simulation experience. Furthermore, the measure had to be able to independently detect the existence of change along any domain. Also, since the magnitude of the change to be measured was not known, the instrument had to be sensitive enough to signal small changes should they occur. Another instrument characteristic suggested by the findings of Robinson (1967), Alger (1967), and Rossi et. al. (1967), was indirectness, to avoid the tendancy of faking responses.

Thus, a summary of the essential characteristics, abstracted from the review, of the instrument used to measure the outputs of simulation yielded three main points. The instrument must measure change in differing domains. It must be able to detect the range from small to large and specific--non-specific changes. The third characteristic was that the instrument be of an indirect, non-obvious type.

Instrument selection. A review of the traditional sources as well as some newer textbooks (Fox, 1969; McGrath, 1970; and Englehart, 1972) suggested several techniques, the most promising of which, because of its flexibility and the amount of supporting research, appeared to be the semantic differential technique. Support for this particular choice can be drawn from the study by Bell (1971) who used the semantic differential for similar purposes, namely, the measurement of change caused by simulation, and from Brod, Kernoff, & Terwilliger, (1969), who used a semantic differential instrument in situations with a high level of anxiety.

The methodology of the technique is extensively detailed by Osgood et. al., (1967), in their book The Measurement of Meaning, and therefore will not be dealt with here. Osgood's technique permitted a wide range of concepts to be used. To maintain the broad perspective that was so essential to this initial investigation it appeared that a minimum of two concept areas must be measured. These two areas contained the majority of the participant's experiences that occurred during this study.

The first area pertained to the technique of simulation. A semantic differential instrument using the concept of simulation techniques for educational purposes had the subjects describe the meaning of this concept before and after the treatment. However, measuring the participant's meaning of simulation was not enough. It seemed reasonable to maintain that the participant's view of the content presented by the simulation might have changed as a result of his being exposed to it while experiencing some stress and anxiety. Since the concept of simulation techniques for education purposes was pertinent only to the technique employed, a measure of content was also required. Thus a second semantic differential instrument was developed to determine any change in the meaning of the content presented during the simulation.

Instrument development. A pilot version of the instrument was initially developed using both concept and scale selection procedures suggested by Osgood. The two main criteria involved factorial composition and relevance. For the pilot instrument relevance was judged using the investigator's subjective opinion. Initial factorial composition was determined from a review of various studies selected from Osgood's book. In addition several adjective pairs (scales) were developed by the investigator specifically for this study. The end result was six concept semantic differential instrument encompassing the two major areas of investigation, each at three levels of specificity. Thirty six scales were used and were systematically distributed with respect to the concepts and origin of the scale.

The pilot instrument was administered to a group of 118 students with similar characteristics to those of the sample. The results of this administration were factor analyzed using the Biomed X72 program (Varimax

Rotation). The resulting factor loadings served as the basis for making the choice as to the 24 scales used in the final experimental instrument.

Reliability of the output measures. Split-half reliability of the pilot instrument was calculated by randomly dividing the scales loading on specific factors into two equal groups and then correlating the scores on these scales for each half. This procedure was repeated for each major set of concepts, simulation and conflict, and for each of the first three factors. Because the split-half reliability method typically results in a low estimate (Fox, 1969) the estimated reliability was adjusted using the Spearman-Brown prophecy formula. Table 1 presents the results of these calculations. Very clearly, the desired and necessary reliability was achieved.

Validity of output measures. The validity of the semantic differential measures, both pilot and experimental, rests largely on the content validity of the concepts used. A rationale meeting this requirement, has been described in the prior section dealing with the selection of the concepts to be used. Thus, it was held that the instrument possessed the needed degree of content validity.

Data Collection Procedures. The volunteers comprising the sample of this study were asked to schedule a mutually convenient time with the experimenter. Each experimental session, involving only one subject, required approximately one hour. When the subject arrived at the research site, he saw only the outside of the simulator. No equipment of any kind was visible. The investigator attempted to set the subject at ease. The subject was then asked to take the pre-treatment semantic differential instrument after which the electrodes were applied according to the procedures developed by the pilot runs. At this time the subject was reassured that no shock would be given.

Table 1

Pilot semantic differential instrument: Split-half reliability coefficients for factors obtained by collapsing across levels of generality

Test	Factor one[a]		Factor two[a]		Factor three[a]	
	Pearson r	Adj. r[b]	Pearson r	Adj. r[b]	Pearson r	Adj. r[b]
Simulation[c] 01/001 02/001 03/001	.69	.81	.30	.47	.43	.61
Conflict[c] 04/001 05/001 06/001	.65	.79	.70	.83	.62	.76

a. All adjusted correlations are significant beyond .005, one tailed test

b. Adjusted using Spearman-Brown prophecy formula, $r(adj.) = \frac{2r}{1+r}$

c. N=118

Treatment involved the experimenter's retreat from the DAS and the closing of the door. The subject was then told that a two minute silence, necessary to establish an initial physiological baseline, would precede the introduction to the simulation. Once the two minute point was reached, the program switched on. Because of the mechanical nature of the program's sequencing mechanism, equivalent treatment was assured.

When the program ended the participant was asked to remain at his desk and relax for an additional two minutes to establish the terminal physiological baseline. The experimenter entered at the end of this period and administered the post-treatment semantic differential instrument. Upon completion, the subject had the electrodes removed.

DATA ANALYSIS

Three sets of data: two sets, one pre and one post treatment, of semantic differential data and one set of physiological data, were available for analysis. The three sets of data were treated by two separate stages of analysis. The first stage was designed to convert raw data to a form that facilitated hypothesis testing and it was this latter aspect that became the second stage of analysis.

Initial Analysis

The response data was subjected to the BMD X 72 factor analysis program. The purpose of this analysis was to (a) verify the scale selection as detailed in the previous chapter and (b) to determine the factor structure of the responses.

Three factors were rotated. The factor analysis identified which scales load on which factors. Depending on the instrument, the first three factors accounted for a minimum of 58% and a maximum of 71% of the total variance.

[illegible] their constituent scales were identified (Table 2) [illegible] each factor was developed. It must be pointed [illegible] alternative descriptions that could be [illegible], [illegible] measuring change the important characteristic [illegible] scales on factors, not the ongoing definition of that [illegible].

The concept of simulation is characterized, for this sample, by Factors A, B, and C. Factor A consists of the scales: good--bad, fair--unfair, real--unreal, important--unimportant, and powerful--powerless. The first three scales comprise an evaluative assessment of the concept. As implied by the last two scales, this assessment is combined with what

Table 2

Experimental semantic differential instrument: Best fitting scales

Tests	Factor A[a]	Factor B	Factor C
07/003 07/004 Simulation	important--unimportant	cold--hot	flexible--rigid
	good--bad	exciteable--calm	changeable--constant
	fair--unfair		
	real--unreal		
	powerless--powerful		

Tests	Factor A[a]	Factor G	Factor H
08/003 08/004 Principalship	important--unimportant	thick--thin	potent--impotent
	real--unreal	changeable--constant	fair--unfair
	active--passive	cold--hot	dishonest--honest
	impressive--unimpressive		
	powerless--powerful		

a. To avoid possible confusion between simulation and principalship, factors were labeled with letters.

appears to be some judgment as to the potency of the technique. Factor B places the meaning of the concept along the dimension of activity. That is to say, "hot" and "exciteable" things are more active than "cold" and "calm" things. In a similar manner, Factor C, consisting of the scales: flexible--rigid and changeable--constant, locates the concept's meaning along a dimension that appears to define an ability to adapt. It may also be that this is a slightly different aspect of the activity dimension.

Principalship as a concept is also defined by three factors, in this case they were designated F, G, and H to eliminate confusion. Factor F, composed of scales: important--unimportant, real--unreal, active--passive, impressive--unimpressive, and powerless--powerful, seems similar to Factor A. Factor F also includes some measure of activity. Thus, this factor encompasses all three dimensions and thereby it is not as clear and definitive as A. Factor G is a combination of potency: thick--thin, and activity: changeable--constant, cold--hot. Factor H is also a combination of potency: potent--impotent, and evaluation: fair--unfair. From these descriptions, it can be seen that Factors A, B, and C define simulation more cleanly and with less interaction than Factors F, G, and H define principalship.

Since the purpose of this study was to investigate the relationship between concept meaning change and physiological activity, the next step was to quantify the exact change in meaning. A program was constructed that averaged each participant's responses for the scales comprising each factor. In addition, group averages were calculated and are indicated in Table 3. Both pre and post treatment averages were calculated in this way.

A correlated test was used to test the mean of the entire group's difference between pre and post treatment scores. This was repeated for each factor of each concept. Because a difference was expected, a one tailed interpretation was employed. The differences for the three factors describing simulation were significant at the .1 level; however, none of the group differences for the principalship factors could be so interpreted.

Each participant's factor score was also processed by a program that calculated the D statistic:

$$D = \sqrt{(A - A')^2 + (B - B')^2 + (C - C')^2}$$

The D statistic is the distance between two points in a three dimensional euclidian space. The first point was located by using the pretreatment factor means as coordinates and the second was determined by using the post treatment factor means in a similar manner. The D statistic thus calculated formed the index of change correlated with physiological activity.

Table 3

Experimental Semantic Differential Instrument:
Group Mean Factor Scores

Subject	Factor	Simulation		Factor	Principalship	
		Pre	Post		Pre	Post
entire	A*	2.44	2.31	F	2.21	2.20
group	B*	3.00	3.45	G	3.50	3.72
mean	C*	2.74	3.24	H	2.18	2.05

* Significant difference. .1 level, one tailed correlated t test

Physiological Activity

The physiological measures recorded were HR and GSP activity (not level). The type of recording method allowed only frequency measures for HR, but both amplitude and frequency could be obtained from the GSP recording. In addition to the recording during the simulation, two one minute baselines, one before and one after treatment, were also taken. The baselines permitted calculation of deviation scores as recommended by Burstein et. al., (1965). These scores reflect the difference between a participant's resting score and his scores during each minute of simulation. Because deviation scores calculated in this way compensate for different reaction sensitivities, it was assumed that such scores are comparable across individuals.

General scoring procedure. The simulation program was structured so as to allow varied response times. This resulted in varying total simulation entry-exit times. Accordingly, to equalize the situation for statistical treatment, the scores were all prorated to the most common duration, namely 19 minutes. Furthermore, since all motion artifact could not be controlled by electrode placement, sections of the physiological traces were discarded wherever motion artifact was detected. The remaining portion of the useful trace was further prorated in proportion to the removed portion.

Heart rate. The simplest of the recorded indices, heart rate, was scored by counting the number of r-spikes in each minute.

GSP. This measure was scored to provide two indices of activity, amplitude and frequency. The frequency index was defined as the number of GSP events that occur within a given time. Every trace peak (event) contained between two points of inflection, either above or below the trace centerline, was counted to yield this score. This method of scoring was called the individual method.

The project consultant also suggested an alternative scoring procedure for the frequency index. This method consisted of analyzing the trace to determine the number of complete cycles occurring per minute. This scoring approach was called the cyclic method. Because of the possibility that mono-, bi-, and tri-phasic responses could occur within the same trace and the subsequently difficulty in determining which event had which characteristic, it seemed that this method was more subjective than the individual method. However, because the consultant pointed out the possibility of uncovering relationships due to the non-linearity of adjustment, his advice was heeded. Consequently, both methods of frequency analysis were used.

The amplitude index was defined as the sum of the vertical displacements of each peak from the trace centerline. Note that this index is a total of all deviations regardless of their directions. Burstein et. al., (1965) support this method when they report that "the most effective SP index of stimulus induced emotionality is the sum of all vertical deviations from the baseline (Page 24)."

Interpretation. Several alternative interpretation and scoring procedures were made possible by the methods of data collection. A review of the literature did not yield conclusive evidence, nor even a predominance of opinion as to the recommended procedures, so a deductive approach was taken. It was assumed that the baselines, because they were taken in a no stimuli situation, represent a resting level of the organism's physiological activity. Accordingly, when stimuli are injected, as is the case when the simulation is in progress, the activity level should be changed. Thus the algebraic sum of the difference between each minute's score and the baseline, yields a gross index of the level of activity, which because of the baseline reference is comparable among individuals. However, beside the above, a further possibility is that any deviation, either positive or negative, from the baseline is indicative of activity and that in this event, the algebraic sum would be inappropriate and that instead the absolute sum would represent a more correct index of activity.

Beyond these questions, one further logical issue remains to be resolved. This issue pertains to the use of the baselines. Because two, one before and one after treatment were taken, the two can be averaged to provide a longer period for a baseline; or, it could be argued that since the second baseline occurs immediately after the simulation ends, not enough time has elapsed for the participant to return to a resting state, and that consequently only the first baseline is a valid indicator of the resting state.

To insure a comprehensive analysis that encompassed many of the possibilities of this exploratory investigation, it was decided to use each of the interpretation methods described in the preceding paragraphs. It was also felt that this decision would provide the broadest data base upon which subsequent investigators could build.

Secondary Analyses

The main purpose of this stage was to determine the relationships between physiological activity and change in concept meaning, and then use this to test the hypotheses. A correlation of each subjects' D score with each of the various indices of physiological activity provided the necessary test of the hypotheses as well as providing additional insights as to the effects of simulation. The Pearson Product Moment correlation coefficient was used in all the calculations shown on Table 4.

Once Table 4 became available and the low correlations were noted, sample scattergrams were plotted from the same data that yielded the correlations. These scattergrams, while typical of low correlations, did not suggest the presence of any severe departure from linearity, a necessary condition for correlational analyses. In addition, the assumption of homoscedasticity was reviewed, but because of the small sample it was difficult to determine the actual condition of this characteristic. Because of the analytical techniques employed, it was assumed that the condition was met to a satisfactory degree.

While some of the correlations reported in Table 4 are significant, these coefficients represent only a relationship between a single physiological variable and change in concept meaning. As indicated by the review, combinations of physiological variables however, provide a more inclusive representation of involvement or the degree of activation. Consequently, the coefficient of multiple correlation (R) (Tate, 1959, p. 309) was calculated. The results are reported in Table 5.

The multiple correlation coefficient gives the relationship between two or more predictor variables, taken as a team, and a single criterion variable. This study employed both two and three predictor models of this coefficient. Physiological variables were taken exclusively as predictor variables and change in concept meaning (D) was taken exclusively as a criterion. To determine if simplification was possible, two predictor correlations were calculated and the results presented in Table 6.

FINDING & CONCLUSIONS

Semantic Differential

Instrumentation and technique. Analysis of pre and post treatment semantic differential factor means suggested systematic motion of factor means rather than a random scattering. This tendency combined with the stability of the semantic differential responses supports the finding that the semantic differential technique is indeed a very sensitive indicator of concept meaning. It must be emphasized, however, that this is probably only true when strict developmental procedures and pilot instruments are used. Since high reliability correlations were identified, this allows greater confidence in the D statistic calculated from the factor averages.

Table 4

Correlation matrix:
Physiological indices versus change in concept meaning[a]

Physiological indices	Change in concept meaning (D) Simulation	Principalship
$\Sigma\Delta HR$	-.22	-.07
$\Sigma\Delta HR\bar{x}$	-.17	-.08
$\lvert\Sigma\Delta HR\rvert$	-.16	-.16
$\lvert\Sigma\Delta HR\bar{x}\rvert$	-.13	-.14
$\Sigma\Delta GSPF$	.12	.20
$\Sigma\Delta GSPF\bar{x}$	.24	.14
$\lvert\Sigma\Delta GSPF\rvert$	.04	.13
$\lvert\Sigma\Delta GSPF\bar{x}\rvert$	.13	.17
$\Sigma\Delta MGSPF$	.04	.11
$\Sigma\Delta MGSPF\bar{x}$	.02	.13
$\lvert\Sigma\Delta MGSPF\rvert$	-.12	.01
$\lvert\Sigma\Delta MGSPF\bar{x}\rvert$	-.03	.15
$\Sigma\Delta GSPA$	.24	.18
$\Sigma\Delta GSPA\bar{x}$	.33*	.20
$\lvert\Sigma\Delta GSPA\rvert$	.30*	.18
$\lvert\Sigma\Delta GSPA\bar{x}\rvert$	.37*	.22
$AV.\Delta GSPR$	.09	-.04
$AV.\Delta GSPR\bar{x}$	-.06	.03
$\lvert AV.\Delta GSPR\rvert$	.19	.02
$\lvert AV.\Delta GSPR\bar{x}\rvert$	.28	.06
$AV.\Delta MGSPR$	.22	.07
$AV.\Delta MGSPR\bar{x}$	.58*	-.04
$\lvert AV.\Delta MGSPR\rvert$	.38*	.16
$\lvert AV.\Delta MGSPR\bar{x}\rvert$	.29*	.10

a. N = 32

* Significant, .05, one tailed

Table 5

Three predictor-variable multiple correlations:
Physiological indices predicting change in concept meaning

Physiological indices	Change in concept meaning (D)	
	Simulation	Principalship
$\Sigma\Delta HR$ $\Sigma\Delta GSPA$ $\Sigma\Delta GSPF$	.32	.24
$\Sigma\Delta HR$ $\Sigma\Delta GSPA$ $\Sigma\Delta MGSPF$	.36	.20
$\Sigma\Delta HR\bar{x}$ $\Sigma\Delta GSPA\bar{x}$ $\Sigma\Delta GSPF\bar{x}$	.39	.22
$\Sigma\Delta HR\bar{x}$ $\Sigma\Delta GSPA\bar{x}$ $\Sigma\Delta MGSPF\bar{x}$	.39	.24
$\lvert\Sigma\Delta HR\rvert$ $\lvert\Sigma\Delta GSPA\rvert$ $\lvert\Sigma\Delta GSPF\rvert$	.35	.23
$\lvert\Sigma\Delta HR\rvert$ $\lvert\Sigma\Delta GSPA\rvert$ $\lvert\Sigma\Delta MGSPF\rvert$	.40	.22
$\lvert\Sigma\Delta HR\bar{x}\rvert$ $\lvert\Sigma\Delta GSPA\bar{x}\rvert$ $\lvert\Sigma\Delta GSPF\bar{x}\rvert$	.40	.26
$\lvert\Sigma\Delta HR\bar{x}\rvert$ $\lvert\Sigma\Delta GSPA\bar{x}\rvert$ $\lvert\Sigma\Delta MGSPF\bar{x}\rvert$	.40	.28

All multiple correlations failed to attain a significant $F_{3,28}$ value, $P<.1$

Effect of the simulation. It is quite clear that the semantic differential instrument dealing with the technique of simulation was more effective than the instrument pertaining to the principalship. It was determined that the former accounts for more variance attributable to the three factors than the latter. This finding has the effect of weakening the usefulness of criterion scores calculated from the principalship instrument.

The pre and post treatment semantic differential instruments dealing with simulation indicated an average D statistic of .69 units. Figure 2 shows the dimensional components of this value and graphically illustrates the shift in meaning that resulted from treatment. Factor A, indicative of the evaluative-potency dimension, changed from 2.44 to 2.3 or .13 units towards a more positive meaning. Factor B, an index of the activity dimension, changed from 3.00 to 3.45 or .45 units towards a less active meaning of simulation. Factor C, an index of adaptability, has moved .50 units towards a less flexible of adaptable meaning.

Analysis of the pre and post treatment semantic differential instruments pertaining to the principalship yielded an average D statistic of .25 units. This is a very small and perhaps even insignificant shift. When this statistic's components were analyzed, no single dimension was found to change more than .224 units. Therefore, it appeared reasonable to accept the finding that no change in concept meaning took place. The graphic representation of the principalship's D statistic's dimensional components is presented in Figure 3.

If, as is suggested by the above paragraph, there is no shift in what the concept of principalship means to the participants, all correlations of physiological variables with concept meaning change must be insignificant. Tables 4, 5, and 6 support this finding.

Table 6

Two predictor-variable multiple correlations:
Physiological indices predicting change in concept meaning

Physiological indices	Change in concept meaning (D)	
	Simulation	Principalship
$\Sigma\Delta HR$ $AV.\Delta GSPR$	.23	.08
$\Sigma\Delta HR$ $AV.\Delta MGSPR$	.30	.10
$\Sigma\Delta HR\bar{x}$ $AV.\Delta GSPR\bar{x}$	.18	.08
$\Sigma\Delta HR\bar{x}$ $AV.\Delta MGSPR\bar{x}$	.59*	.09
$\lvert\Sigma\Delta HR\rvert$ $\lvert AV.\Delta GSPR\rvert$	.23	.16
$\lvert\Sigma\Delta HR\rvert$ $\lvert AV.\Delta MGSPR\rvert$	.40	.21
$\lvert\Sigma\Delta HR\bar{x}\rvert$ $\lvert AV.\Delta GSPR\bar{x}\rvert$	.29	.15
$\lvert\Sigma\Delta HR\bar{x}\rvert$ $\lvert AV.\Delta MGSPR\bar{x}\rvert$	.31	.17

All multiple correlations, except *, failed to attain a significant $F_{2,29}$ value, $P<.1$

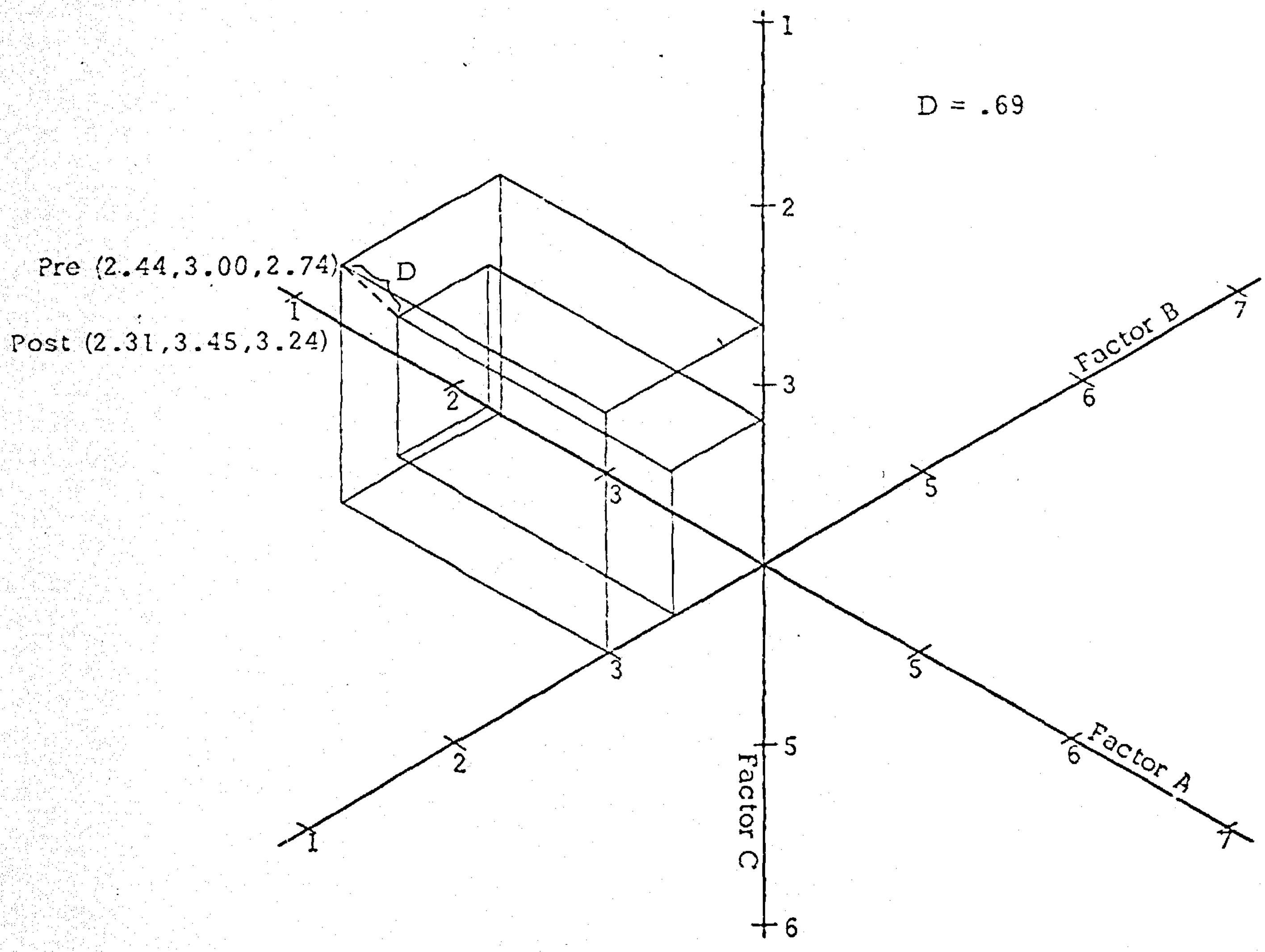

Figure 6. Graphic representation of D components for simulation concept

Both D statistics, indicative of shift in concept meaning, share the characteristic of changing most in the second and third factors. The first factor of each concept appears not to change very much if at all. This finding would suggest that the evaluative aspects of a concept's meaning are not as easily changed as are those aspects involving activity and potency.

Physiological Indices

By not evidencing any repetitive pattern, for example, not all deviation scores calculated from the first baseline correlating highest, Table 4 exposed the finding that the various methods of scoring and interpretation do not greatly vary the resulting physiological--concept-meaning change correlations. No evidence was found to favor the use of either the algebraic or absolute method of summing deviation scores. Using similar reasoning, there is no evidence supporting or rejecting either the single baseline or the pre and post treatment baseline method of obtaining deviation scores. Using similar reasoning, there is no evidence supporting or rejecting either the single baseline or the pre and post treatment baseline method of obtaining deviation scores. However, if the post treatment baseline is taken shortly after the simulation ends, it is cautioned that this finding may only be true for those simulations which do not raise the anxiety level of the participants to an excessive degree.

Relationships between Semantic Differential and Physiological Indices

Strength, linearity, and direction. The relationships described by the multiple correlations in Tables 5 and 6 are all positive as they of course must be. In addition, all significant single physiological variable--concept-meaning change based D statistic correlations, provided in Table 4, are positive. Generally, however, all single physiological variable based

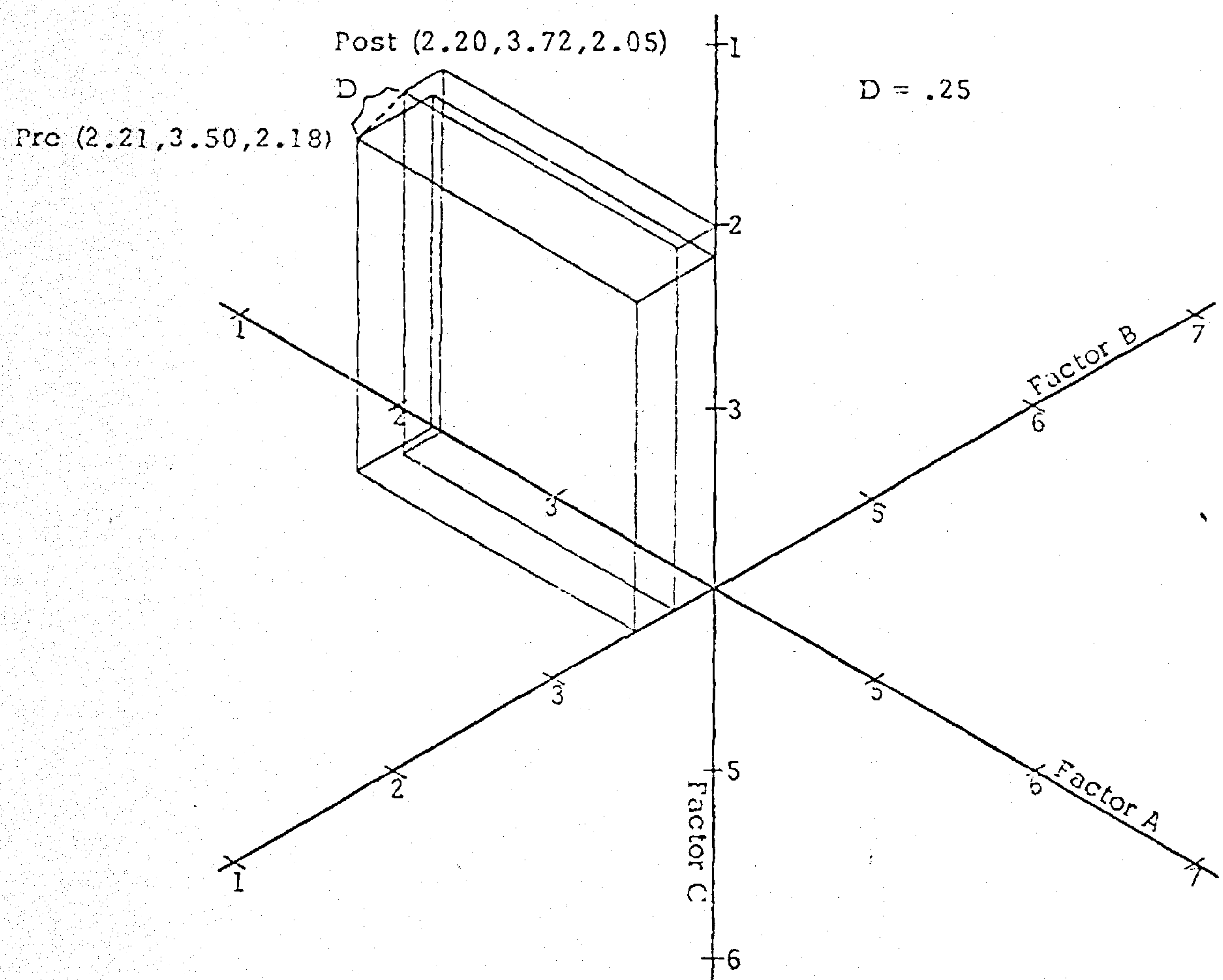

Figure 7. Graphic representation of D component for principalship concept

correlations are very weak. Correlations based on the average GSP response, that is the bottom four in Table 4, are distinctly stronger than either frequency or amplitude correlations alone.

The multiple correlations reported in Table 5 are higher and thus account for more variance than single physiological variables. Even though this is the case, the multiple correlations evidenced in the data generated by this study are interpreted to be weak.

While the individual indices of heart rate, GSPA, and GSPF do not correlate well with the criterion, a derivative from these scores does. When GSPR, which is a rough estimate of the average GSP response, is calculated, it quite clearly correlates better than any other single variable. It must be pointed out that this is only the case when frequency is calculated by the cyclic method as was suggested by the project consultant.

A comparison of Tables 5 and 6 results in the finding that with one exception in 16 cases, multiple correlations using three predictor variables are higher and thereby account for more variance than multiple correlations with two predictors. It must also be noted that in the two predictor case, one of the two predictor variables was GSPR, which is, in fact, a composite indicator resulting from the division of GSPA by GSPF. This results in the finding that more variance is accounted for when combining three variables individually than by working with a subgrouping of them.

The multiple correlations, in both three-and two-predictor forms, were always bound to be higher when employing data generated using the cyclic method of frequency scoring.

Conclusions

Because this study employed correlation analysis, causal statements as to the effect of the simulation treatment are not permissible. Beyond this limitation, however, correlations impose the requirement of interpretation of the degree of relationship detected. Categorical claims cannot be supported with these statistics. The following conclusions are offered within these restrictions.

The study's general hypothesis appears to be acceptable. That is, for this particular sample, a positive relationship between the degree of involvement as indicated by physiological activity, and change in concept meaning was evidenced. The degree of strength of this relationship is interpretated to be moderate, but systematically present. Accordingly the strength discovered was deemed to be high enough to accept the hypothesis and to warrant use in the other investigations contemplated for Project CORES.

The multiple correlations clearly support the conclusion that a number of predictors are of greater value than any single predictor variable. The predictor selection procedure as well as the final correlations both support the conclusion that the three physiological variables used in this study represent a minimum treatment of covert responses to simulation. However, the small amount of variance (R^2) accounted for by the three predictor variables clearly suggests the conclusion that other significant factors of influences on simulation remain to be identified.

The findings allow the conclusion that differing methods of interpretation, scoring, and summing of deviations scores may provide meaningful information. This same data also requires the conclusion that the cyclic method of scoring frequency is more useful than the individual method.

BIBLIOGRAPHY

Alger, R. In P. H. Rossi & B. J. Biddle (Eds.). The new media and education: Their impact on society. Garden City, N. Y.: Doubleday, 1967.

Beck, I. H., & Monroe, B. Glossary. Educational technology, 1969a, (10), 44.

Bell, C. G. An assessment of simulation training on attitude change of prospective teachers. Paper presented at the Annual Meeting of the American Educational Research Association, New York, 1971.

Brod, D., Kernoff, P., & Terwilliger, R. Anxiety and semantic differential responses. In J. G. Snider & C. E. Osgood (Eds.), Semantic differential technique: A source book. Chicago: Aldine, 1969.

Burstein, K. R., Fenz, W. D., Bergeron, J., & Epstein, S. A comparison of skin potential and skin resistance as measures of emotional responsivity. Psychophysiology, 1965, 2(1), 14-24.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: McNally, 1963.

Cherryholmes, C. H. Some current research on effectiveness of educational simulation: Implications for alternate strategies. American behavorial scientist, 1966, 10(2), 4-7.

Cruickshank, D. E., & Broadbent, F. W. The simulation and analysis of problems of beginning teachers. U.S.O.E. cooperative research project No. 5-0789 Washington, D.C.: United States Government Printing Office, 1968.

Dyrenfurth, Michael. Investigation of relationship between participant involvement in stimulation as evidence by physiological activity, and change in concept meaning. (Doctoral dissertation, Bowling Green State University) Ann Arbor: 1973. No. 1523-a.

Englehart, M. D. Methods of educational research. Chicago: McNally, 1972.

Fox, D. J. Research process in education. Toronto: Holt, Rinehart-Winston, 1969.

McGrath, J. H. Research methods and designs for education. Scranton, Pa.: International textbook, 1970.

Osgood, C. D., Succi, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1967.

Robinson, J. A. Simulation and games. In P. H. Rossi & B. J. Biddle (Eds.), The New media and education: Their impact on society. Garden City, N. Y.: Doubleday, 1967.

Rossi, P. H. & Biddle, B. J. (Eds.) The new media and education: Their impact on society. Garden City, N. Y.: Doubleday, 1967.

Tate, M. W. Statistics in education. New York: Macmillan, 1955.

Twelker, P. A. Designing simulation systems. Educational Technology, 1969, 9(10), 64-70.