Animal Learning & Behavior
1985, 13 (4), 339-344

The principle of aggregation in psychobiological correlational research: An example from the open-field test

KLAUS-PETER OSSENKOPP and DWIGHT S. MAZMANIAN
University of Western Ontario, London, Ontario, Canada

The principle of aggregation states that the sum of a set of multiple measurements is a more stable and representative estimator than any single measurement. This greater representation occurs because there is inevitably some error associated with measurement. By combining numerous exemplars, such errors of measurement are averaged out, leaving a clearer view of underlying relationships. The present study explored the effect of score aggregation over various time periods on correlations among a number of reliable measures frequently used in open-field testing. Twenty-six male rats were given four open-field tests (4 min in duration) at 48-h intervals. Ambulation, rearing, and defecation responses were measured on a minute-by-minute basis in the open-field tests. Correlation matrices were calculated among the three measures for unaggregated scores (1-min totals) and for scores aggregated over daily tests, and mean correlation coefficients were computed for all three pairwise comparisons of the three response variables. These mean correlations were then compared to those obtained when the open-field measures were aggregated over all 4 test days. The results showed that aggregation produced substantial increases in correlation-coefficient magnitude. The correlation between ambulation and rearing increased from a mean of .39 to a value of .81. Similar increases were observed when defecation scores were correlated with ambulation (−.17 to −.59) and rearing (−.16 to −.49). Thus aggregation is an important factor to be considered in the design of psychobiological correlational studies.

In recent years, the principle of aggregation has received renewed interest in the areas of personality (Epstein, 1979, 1980) and behavioral development (Rushton, Brainerd, & Pressley, 1983). This principle is based upon the notion that the sum of a set of multiple measurements is a more stable and unbiased estimator than any single measurement from the set. Because any measurement has an error component associated with it (cf. Gulliksen, 1950), the combining of several measurements tends to average out these error components, providing a better estimate of the true value of the parameter in the population of interest. Perhaps the best-known example of this principle is the common rule that the larger the sample of a set of measurements, the more representative is the sample mean of the population mean.

In psychobiological research, as in personality or developmental research, the principle of aggregation can be used to obtain better estimates of the true values of various behavioral, physiological, or neurochemical parameters. A well-known rule in educational and personality testing is that the reliability of an instrument increases as the number of test items increases (e.g., Gulliksen, 1950; Lord & Novick, 1968). Another example of this principle is the common practice of averaging the decisions of several judges to obtain relative rankings of subjective qualities of people of their behavior in various kinds of competitions, such as those related to physical beauty, artistic ability, or athletic ability.

Statisticians have often made the point that several measurements of a phenomenon are better than one. An important example comes from an early paper by Spearman (1910) on the proper use of correlation coefficients:

> It is the superposed accident (measurement error) that the present paper attempts to eliminate, herein following the custom of all sciences, one that appears to be an indispensable preliminary to getting at nature's laws. This elimination of the accidents is quite analogous to, and serves just the same purpose as, the ordinary process of "taking means" or "smoothing curves."... The method is as follows. Let each individual be measured several times with regard to any characteristic to be compared with another. (pp. 273-274)

In fact, it is well recognized that a reasonable number of subjects need to be sampled in order to obtain meaningful estimates for any given population parameter—few people would consider an experiment with one subject per experimental condition as adequate. However, aggregation has been widely ignored with respect to the sampling of stimuli and occasions in laboratory studies.

Epstein (1980) identified four forms of aggregation: over subjects, over time, over stimulus situations, and over modes of measurement. Aggregation over subjects

This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (U0151) to the first author. We thank J. P. Rushton for helpful discussion. Please send reprint requests to: K.-P. Ossenkopp, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

has already been mentioned; however, because it is well known and is dealt with in all introductory statistics courses, it can be used as an instructive example (cf. Epstein, 1980; Gulliksen, 1950). It is well known that the larger the size of a sample drawn from a population of interest, the more representative the sample mean is of the population mean. The random errors associated with each measurement tend to average out and the resultant sample statistic will more closely approximate the population parameter. Increasing sample size also has the effect of increasing generalizability from the sample to the population (Gulliksen, 1950). Aggregating over trials and/or occasions cancels out the uniqueness of particular trials and/or occasions, and thus increases temporal reliability or replicability (Epstein, 1980). However, data derived from a subject's first encounter with a stimulus situation may differ considerably from data obtained on subsequent tests, and experiments that rely on single-occasion testing may not be generalizable to multiple-occasion situations (see Epstein, 1979, 1980). Aggregating over stimulus situations cancels out the unique effects associated with particular stimulus situations, and thus allows generalization to a larger domain of stimuli. However, in many studies, fairly restricted ranges of particular stimulus values are of focal interest, and only limited aggregation over these stimuli is possible. The effects of aggregating over time (trials and/or occasions) and over stimulus situations need to be investigated empirically in psychobiological research.

When Rushton et al. (1983) applied the principle of aggregation to correlational research studies in behavioral development, they found that substantial increases in correlations between measures could be obtained by aggregating over several measures. In psychobiological research, a single behavioral measure is often correlated with a single physiological or biochemical measure (see review by Will, 1977), with resultant small or negligible correlations. Will (1977) recognized that aggregation was important for correlational analysis by suggesting that multiple measures of the variables of interest be obtained:

> The biochemical measurement can be done on such small quantities of substrate that it can be repeated, generally, on randomized fractions of the sample. In most cases, the behavioral test can also be repeated, although it is difficult to maintain the individuals' constancy and especially the constancy of their motivational levels. (p. 160)

However, few research studies have in fact employed this approach.

The Open-Field Test

Since its development by Hall (Hall, 1934, 1936; Hall & Ballechey, 1932), the open-field test has attained the status of one of the most widely used instruments in animal behavior analysis (see reviews by Archer, 1973; Walsh & Cummins, 1976). Typically, the open-field apparatus consists of a novel open space from which escape is prevented by a surrounding wall. An animal, usually a rodent such as a rat, is placed in this apparatus for some fixed time interval and the incidence (and, if applicable, the duration) of certain behaviors is recorded. The number of different behaviors quantified in the open-field test has increased to over 20 (cf. Walsh & Cummins, 1976), but only a few have been shown to be reliable for the rat. Ambulation (number of subdivisions entered), rearing frequency, and defecation (number of fecal boli deposited) have been found to be reliable open-field measures (Ivinskis, 1968), and also tend to be the most common measures used.

Correlations among open-field measures have been of interest because some have been viewed as measures of more general constructs, such as emotionality (defecation) or exploration (ambulation, rearing), and because several parameters might measure the same psychophysiological state. Correlations between ambulation and defecation scores have been studied most often, and there seems to be general agreement that a negative correlation exists between these two measures, although some studies have failed to find significant correlations (see Archer, 1973, Table V, for a review of the rat literature). Correlations between ambulation and rearing, not surprisingly, have usually resulted in significant positive correlations (e.g., Anderson, 1938; Ivinskis, 1968; Pare, 1964; Ray & Hockhauser, 1969), even though these two motor behaviors are mutually exclusive. When rearing and defecation measures have been correlated, negative correlations have usually been found, but these are of smaller magnitude than those between ambulation and defecation (e.g., Holland & Gupta, 1966; Tachibana, 1982).

In the present study, our intention was to determine what effect aggregation of open-field scores over minutes of testing and over test days would have on the intercorrelations within and across three reliable open-field measures (ambulation, rearing, and defecation). We were also interested in the effects of score aggregation on correlations between motor behaviors in the open-field test and motor behaviors measured in a stabilimeter apparatus, as well as in a comparison of defecation responses in the two different test situations.

METHOD

Subjects

Twenty-six male hooded rats (Long-Evans strain, obtained from Charles River, Quebec) weighing between 200 and 230 g were used in this study. They were individually housed in stainless steel cages in a colony room with a temperature of about 22° C and were on a 12 h light/12 h dark cycle, with lights on from 0700 to 1900 h. Food and water were available ad lib.

Apparatus

Open field. The open field was circular, with a diameter of 90 cm, and was enclosed by a wall 30 cm high. The floor and wall of the open field were black and the floor was divided into 25 equal area sections by thin white lines. The floor was covered with a transparent plastic coating. The open-field apparatus was located inside a large wooden frame which was surrounded on all sides by black curtains. The field was illuminated by two 60 W fluorescent lights located 100 cm above the field. A white-noise generator provided a masking noise of 61±1 dB at the floor of the open field (measured with a Bruel & Kjaer Sound-level meter, type 2203).

Stabilimeter. Two animal activity monitors (Lafayette, Model A-501) were used to measure general activity. These instruments transduce movement of the animals to electrical signals by means of a magnet and coil. The electrical signals are detected by a voltage-sensitive relay which quantifies the signals by activating a counter. Each activity box measured 30×30×30 cm. Electronic programming equipment controlled the data collection and timing.

Procedure

Behavioral testing. Four 4-min open-field tests were administered to each animal, with 48 h separating each test session. At the start of each session the rat was placed in one of the peripheral sections of the open field and the animal's behavior was monitored by an experienced observer for 4 consecutive minutes. The following variables were recorded for each minute of the test session: (1) ambulation—the number of open-field sections crossed by the subject; (2) rearing responses—the number of times the animal raised both forepaws off the floor and extended its body; (3) defecation—the number of fecal boli deposited in the open field. At the end of each session the animal was returned to its home cage and the floor of the open field was cleaned and sponged with a weak vinegar solution to remove any residual odors.

Two days after the last open-field test, and again 48 h later, all animals were tested in the stabilimeter apparatus for 30 min. The number of activity counts and the number of fecal boli deposited in the stabilimeter apparatus during each test session were recorded.

Data analysis. A data matrix was constructed of open-field ambulation, rearing, and defecation scores for each minute of each test session, scores aggregated for each daily test session, scores aggregated over 2-day blocks (Days 1+3 and Days 2+4), and total scores for the four test sessions. Total stabilimeter activity counts and number of defecation responses for each activity test session, as well as the total scores for the two test sessions, were also added to the data matrix. A zero-order product-moment correlation matrix was computed. Mean correlation coefficients were calculated between each pair of open-field variables, using scores for each of the 16 minutes of the open-field tests. Mean correlation coefficients were also calculated using total scores for each of the four daily sessions of the open-field test and scores totaled over 2-day blocks. Mean correlations within each open-field variable were calculated using scores correlated on a minute-by-minute basis and on a day-by-day basis. Finally, mean correlations were calculated for stabilimeter activity and open-field ambulation and rearing, as well as for stabilimeter defecation and open-field defecation. Daily total scores were used to obtain correlations between the stabilimeter test scores and the open-field variables.

RESULTS

The means and ranges of the correlations within each open-field variable, calculated on a minute-by-minute basis, on a day-by-day basis, and on the basis of 2-day blocks, are presented in Table 1. This table shows that scores aggregated into daily totals produced larger intercorrelations and smaller ranges in coefficient values than scores correlated on a minute-by-minute basis. Correlations over split-halves (2-day blocks) produced the largest correlations. The increase in mean correlation value was especially dramatic for the defecation measure, which increased from .11 to .65. Table 2 shows means and ranges of correlations among the open-field variables, calculated on a minute-by-minute, day-by-day, split-halves, and total-score basis. For all three variables, the mean correlation values increased when scores were aggregated into daily totals and the correlations for the total scores gave the largest values. All three of the correlations for the total scores were highly significant (ps < .01).

Table 1
Mean Intercorrelations for Each Variable Over Minutes (Unaggregated), Days (Aggregated), and 2-Day Blocks (Split-Half, Days 1 & 3 and Days 2 & 4) of Testing for the Three Open-Field Measures

Dependent Variable	Minutes		Days		Split-Half
	$\bar{r}$	Range	$\bar{r}$	Range	r
Ambulation	.43	−.08 to .69	.62	.38 to .82	.87*
Rearing	.60	.23 to .82	.76	.68 to .88	.92*
Defecation	.11	−.24 to .66	.53	.28 to .70	.65*

*$p < .01$.

An examination of the day-by-day correlations ($\bar{r}$s) between ambulation and defecation (see Table 3) showed that defecation scores on Day 1 did not correlate very highly with any of the daily ambulation scores ($\bar{r}$=−.14), whereas defecation scores on Days 2, 3, and 4 gave much higher correlations with all daily ambulation scores ($\bar{r}$=−.50, −.43, and −.49 for defecation scores on Days 2, 3, and 4, respectively). A similar pattern was observed for the correlations between defecation and rearing. Day 1 defecation scores correlated, on the average, −.12 with rearing, whereas defecation on Days 2, 3, and 4 correlated, on the average, −.40, −.44, and −.42, respectively. Thus, Day 1 defecation scores followed a pattern different from that of scores on Days 2, 3, and 4 when correlated with ambulation and rearing. This anomaly suggests that perhaps Day 1 defecation scores should not be aggregated with the scores on Days 2, 3, and 4 in calculating overall correlations.

Mean correlations between the open-field measures of ambulation and rearing and the stabilimeter activity measure, as well as those between open-field and stabilimeter defecation scores, are presented in Table 4. The effects of aggregation on these between-apparatus comparisons were similar to the effects on comparisons within the open-field apparatus. Correlations between open-field rearing and stabilimeter activity levels were larger when scores were totally aggregated than when scores were aggregated on a daily basis only. Likewise, defecation levels in the open field and in the stabilimeter test showed a moderate significant correlation only when aggregated over all days. The correlations between open-field ambulation and stabilimeter activity, on the other hand, were not improved by aggregation. The reason for this failure to substantially increase r by aggregation is evident when the data in Table 5 are examined. Open-field ambulation on Day 1 correlated significantly with stabilimeter activity scores on both days, and the correlation was even larger for total stabilimeter activity scores. However, open-field ambulation scores on Days 2, 3, and 4 correlated poorly with stabilimeter activity scores. The data in Table 5 suggest

Table 2
Mean Correlations Among the Three Open-Field Variables Over Minutes (Unaggregated), Days (Aggregated), and 2-Day Blocks (Split-Half, Days 1 & 3 and Days 2 & 4), in Comparison With Correlations for the Final Total Scores

Variables Correlated	Ambulation & Rearing		Ambulation & Defecation		Rearing & Defecation	
Aggregation Level	$\bar{r}$	Range	$\bar{r}$	Range	$\bar{r}$	Range
Minutes	.39	.20 to .60	−.17	.01 to −.31	−.16	−.01 to −.32
Days	.62	.30 to .86	−.39	.01 to −.56	−.35	−.01 to −.63
Split-Half	.77	.70 to .81	−.50	−.36 to −.61	−.44	−.34 to −.52
Total	.81*		−.59*		−.49*	

*$p < .01$.

that aggregation over all 4 open-field test days is not useful in correlating ambulation scores with stabilimeter activity scores. In fact, the relationship evident between Day 1 ambulation and stabilimeter activity is lost when ambulation scores are aggregated over all days. Finally, Table 6 presents the results of applying the principle of aggregation to random data. Random numbers (eight arrays, n=26) were subjected to an aggregation procedure similar to that used for the open-field data. The effect of aggregation was found to be minimal, as should be the case for unrelated data.

DISCUSSION

The results of this study show that aggregating scores over time in the open-field test can result in substantial and significant correlations among the open-field measures of ambulation, rearing, and defecation responses. In general, the greater the degree of aggregation, the larger the mean correlation coefficients obtained among these variables. A similar effect was observed when correlations within variables were calculated. All three open-field measures exhibited substantial mean daily within-variable correlations when scores were aggregated over 2-day test blocks. Finally, aggregation also improved the size of the correlation coefficients obtained between open-field measures and related measures in a stabilimeter activity test.

It is interesting to note that the largest increase in correlation magnitude occurred for the defecation measure (Table 1) when the minute-by-minute mean correlation was compared with the day-by-day mean correlation—an increase from .11 to .53. The low minute-by-minute correlations for this measure may reflect the fact that on a minute-by-minute basis (within a test session), defecation scores are probably not independent. It seems reasonable to assume that once an animal has defecated, the probability of that response subsequently decreases, if for no other reason than the limited supply of fecal boli. If this assumption is true, the result should be negatively correlated defecation scores, because high scores in one minute will be followed by low scores in the next minute. However, on a day-by-day basis, the assumption of nonindependence for defecation scores would not be reasonable, and this should lead to large positive correlations for this measure, if it is a reliable measure. The data in Table 1 (as well as the defecation-related correlations in Table 2) seem to be consistent with the above analysis, and indeed suggest that the defecation measure is a reliable one.

Our observation of a significant positive correlation (.81) between open-field ambulation and rearing is in agreement with the results of a number of previous studies (e.g., Ivinskis, 1968; Ray & Hockhauser, 1969). In the present study, open-field defecation was observed to have a significant negative correlation with ambulation (−.59) and with rearing (−.49), a finding which supports a number of previous studies in which negative correlations were found between defecation and the open-field activity variables (e.g., Hall, 1936; Holland & Gupta, 1966; Tachibana, 1982; Whimbey & Denenberg, 1967). Studies in which negative correlations were not found for these variables may have suffered from too low a level of aggregation for the variables being correlated. A large effect of sampling error might mask any real correlation that exists. However, the relationship between ambulation and defecation seems to be a complex one (see Archer, 1973), and a number of factors other than the aggregation effect are probably important in defining the real relationship between these two variables.

When open-field activity measures of ambulation and rearing have been compared to activity measures obtained in other types of apparatus, some studies have obtained significant positive correlations and others have found

Table 3
Correlations Between Open-Field Defecation and Ambulation and Rearing, Aggregated for Each Day

Variables Correlated		Defecation			
	Day	1	2	3	4
Ambulation	1	−.17	−.43	−.37	−.44
	2	.01	−.49	−.38	−.46
	3	−.13	−.56	−.51	−.49
	4	−.25	−.53	−.45	−.56
Rearing	1	−.05	−.26	−.31	−.32
	2	−.01	−.39	−.36	−.31
	3	−.18	−.55	−.63	−.53
	4	−.23	−.41	−.46	−.52

Note—*Critical r for $p < .05$ is .388, df=24.*

Table 4
Mean Correlations Between Open-Field Variables and Measures Obtained in the Stabilimeter Test

Variables Correlated	Days		Total
	$\bar{r}$	Range	r
Ambulation & Stab. Activity	.24	−.05 to .51	.27
Rearing & Stab. Activity	.34	.20 to .46	.41*
OF Defecation & Stab. Defecation	.28	.12 to .50	.43*

Note—*Open-field (OF) and stabilimeter (Stab.) measures are daily aggregates.* $^*p < .05$.

very low or no correlations (see Walsh & Cummins, 1976). In the present study, we found that open-field rearing (aggregated scores) correlated significantly with activity in the stabilimeter test (.41), but that open-field ambulation did not. These results agree with those obtained by Holland and Gupta (1966), who also found a significant correlation between open-field rearing and stabilimeter activity but not between open-field ambulation and the stabilimeter measure. Our finding of a significant correlation between open-field defecation and defecation in the stabilimeter test ($r=.43$, $p < .05$) is also in agreement with the findings of Holland and Gupta.

At first glance, the failure to obtain a significant correlation between open-field ambulation and stabilimeter activity may suggest that these measures are not related. However, a closer examination of the correlations between daily open-field ambulation scores and daily stabilimeter test scores reveals a more complex situation. Open-field ambulation on the first test day correlated significantly with activity on both stabilimeter test days, whereas open-field ambulation on Days 2, 3, and 4 did not (see Table 5). This observation suggests that open-field ambulation may not be a unitary measure over the 4 test days—that is, that Day 1 ambulation may represent a factor different from that represented by ambulation on Days 2, 3, and 4. Although the positive correlations obtained for the Day 1 ambulation data may represent a Type I error, previous research has indicated that Day 1 ambulation may be qualitatively different from ambulation on subsequent days of open-field testing (Whimbey & Denenberg, 1967). Whimbey and Denenberg found low correlations on Day 1 ambulation with ambulation on subsequent days (a finding also observed in the present study), as well as positive, rather than negative, correlations between Day 1 ambulation and all defecation scores. Ambulation on Days 2, 3, and 4 showed significant negative correlations with all defecation scores. In the present study, Day 1 defecation scores were found to have smaller negative correlations with ambulation and rearing on all days than defecation scores on Days 2, 3, and 4. In general, these observations suggest that Day 1 open-field ambulation and/or defecation may be qualitatively different from these behaviors on subsequent test days—a result supported by a previous study (Tachibana, 1982). Since data derived from a subject's first-time encounter with a stimulus situation may differ considerably from data obtained on subsequent tests (e.g., Epstein, 1979), it may be useful to search for underlying factors that could account for such an effect. Application of a factor analytic technique for longitudinal data (see Harshman, 1970) to the open-field test data showed that Day 1 defecation and urination levels, as well as latency to enter the center area of the field, reflected an emotional reactivity factor. The amount of center-area activity on Days 3 and 4 reflected an exploration factor (Ossenkopp, Sorensen, & Mazmanian, 1986). Other studies have reported similar findings (Walsh & Cummins, 1976; Whimbey & Denenberg, 1967). Such underlying factors could account for the pattern of correlations obtained for the defecation scores over days in the present study. High defecation levels on Day 1 together with moderate activity levels could result in small correlations. Low defecation levels coupled with high activity levels on Days 3 and 4 could result in substantial negative correlations. Such patterns were observed in the present study.

It is also interesting to note that aggregation failed to increase the magnitude of correlations for variables which were not related (e.g., open-field ambulation on Days 2, 3, and 4 and stabilimeter activity, as well as variables created with random numbers; see Table 6). Because aggregation only increases the estimation of the true population parameters of variables, it should not create artificial correlations for variables which are not related. Indeed, other measures collected on the animals used in the present study (see Ossenkopp & Mazmanian, 1985b), measures which did not correlate with the open-field meas ures, did not show an increase in mean correlation magnitude when the variables were aggregated.

Finally, we would like to suggest that application of the principle of aggregation is useful to increase the reliabil-

Table 5
Correlations Between Open-Field Ambulation and Stabilimeter Activity, Aggregated for Each Day

Variable	Day	Activity 1	2	1+2
Ambulation	1	.48*	.51*	.55*
	2	.15	−.05	.07
	3	.24	.09	.19
	4	.30	.21	.29

$^*p < .02$.

Table 6
Correlations Between Eight Arrays (n=26) of Random Digits Unaggregated, Aggregated Over Two Arrays (Split-Half), and Aggregated Over Four Arrays (Total)

Arrays Correlated	$\bar{r}$	Range of r Values
Unaggregated	−.01	−.27 to .39
Split-Halves	−.04	−.19 to .08
Total	.07	

ity (and therefore the generalizability) of measures. These multiple measures can then be examined using a variety of multivariate techniques (e.g., Ossenkopp & Mazmanian, 1985a; Tachibana, 1980). These techniques typically reduce many variables into fewer, more complex variables, which have greater phenomenon realism. Thus, aggregation can be viewed as an important element in psychobiological correlational research.

REFERENCES

ANDERSON, E. E. (1938). The interrelationship of drives in the male albino rat: III. Interrelationships among measures of emotional, sexual, and exploratory behavior. *Journal of Genetic Psychology*, **53**, 335-352.

ARCHER, J. (1973). Tests for emotionality in rats and mice: A review. *Animal Behaviour*, **21**, 205-235.

EPSTEIN, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality & Social Psychology*, **37**, 1097-1126.

EPSTEIN, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist*, **35**, 790-806.

GULLIKSEN, H. (1950). *Theory of mental tests*. New York: Wiley.

HALL, C. S. (1934). Emotional behavior in the rat: Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative Psychology*, **18**, 385-403.

HALL, C. S. (1936). Emotional behavior in the rat: III. The relationship between emotionality and ambulatory activity. *Journal of Comparative Psychology*, **22**, 345-352.

HALL, C. S. & BALLECHEY, E. L. (1932). A study of the rat's behavior in a field: A contribution to method in comparative psychology. *University of California Publication in Psychology*, **6**, 1-12.

HARSHMAN, R. A. (1970). Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-model factor analysis. *UCLA Working Papers in Phonetics* (No. 16), 1-86.

HOLLAND, H. C., & GUPTA, B. D. (1966). Some correlated measures of activity and reactivity in two strains of rats selectively bred for differences in the acquisition of a conditioned avoidance response. *Animal Behaviour*, **14**, 574-580.

IVINSKIS, A. (1968). The reliability of behavioural measures obtained in the open-field. *Australian Journal of Psychology*, **20**, 173-177.

LORD, R. M., & NOVICK, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

OSSENKOPP, K.-P., & MAZMANIAN, D. S. (1985a). The measurement and integration of behavioral variables: Aggregation and complexity as important issues. *Neurobehavioral Toxicology & Teratology*, **7**, 95-100.

OSSENKOPP, K.-P., & MAZMANIAN, D. S. (1985b). Some behavioral factors related to the effects of cold-restraint stress in rats: A factor analytic-multiple regression approach. *Physiology & Behavior*, **34**, 935-941.

OSSENKOPP, K.-P., SORENSEN, L., & MAZMANIAN, D. S. (1986). *Factor analysis of open-field behavior in rats: Using the three-way PARAFAC model for longitudinal data*. Manuscript submitted for publication.

PARE, W. P. (1964). Relationship of various behaviors in the open-field test of emotionality. *Psychological Reports*, **14**, 19-22.

RAY, O., & HOCKHAUSER, S. (1969). Growth hormone and environmental complexity effects on behavior in the rat. *Developmental Psychology*, **1**, 311-317.

RUSHTON, J. P., BRAINERD, C. J., & PRESSLEY, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, **94**, 18-38.

SPEARMAN, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, **3**, 271-295.

TACHIBANA, T. (1980). The open-field test: An approach from multivariate analysis. *Animal Learning & Behavior*, **8**, 465-467.

TACHIBANA, T. (1982). Open-field test for rats: Correlational analysis. *Psychological Reports*, **50**, 899-910.

WALSH, R. N., & CUMMINS, R. A. (1976). The open-field test: A critical review. *Psychological Bulletin*, **83**, 482-504.

WHIMBEY, A. E., & DENENBERG, V. H. (1967). Two independent behavioral dimensions in open-field performance. *Journal of Comparative & Physiological Psychology*, **63**, 500-504.

WILL, B. E. (1977). Neurochemical correlates of individual differences in animal learning capacity. *Behavioral Biology*, **19**, 143-171.

(Manuscript received December 10, 1984;
accepted for publication December 12, 1984.)