

Effects of Participative Decision-Making in Field Hockey Training: A Field Experiment¹

PETER VEEN

Institute of Social Psychology, R. U. Utrecht²

For a considerable time, there has been a conflict, particularly in industrial psychology and organization, about the question of whether democracy is better than autocracy. This discussion is characterized from the one side by a vague definition of the concepts of democracy and autocracy, and from the other side by a great confusion about what would be "better": satisfaction or performance. The vague definition of the concepts is obvious from the multiplicity of the terms used: democracy-autocracy, participation-nonparticipation, consideration-task structuring. Analysis of the literature (Veen, 1970) shows that there are two research traditions.

(1) A series of investigations in the fifties about the dimensions of leadership converged on two independent factors: social-emotional leadership and goal-directed leadership (see, for instance, Bales, 1953; Carter, 1953; Kahn & Katz, 1960). Social-emotional leadership occupies itself primarily with the regulation of the mutual relationships between the members of a group. The goal-directed leadership is primarily directed to attune the human and technical resources in such a way to one another that the task is completed or the group goal is reached. In training both dimensions can be understood as follows: the social-emotional leadership is directed to maintaining the mutual relationship between the members of a training group (the trainer included) at such a level that optimal functioning is possible. The goal-directed leadership consists in giving of exercises, indications, organization of resources, and so on. Next to these two dimensions Carter (1953) indicates a third dimension: individual prominence. This indicates to what extent someone is a strong personality who places himself in the foreground. Mulder *et al.* (1967) have recently investigated this dimension. As this quality seems fairly personality-related, it is of less importance for the present research project.

¹ This research was supported by a grant from the Netherlands organization for pure scientific research (Z.W.O.) to Mauk Mulder, director of the institute, who gave his support and advice to the project.

² Now with W. B. Berenschot Company, management consultants.

(2) Another research tradition focused on participation in decision making (see for instance Coch & French, 1948; Morse & Reimer, 1956). Rather than focusing on what the leadership was aimed at (social-emotional or goal-directed), it focused on the distribution of the leadership over the different persons in a group, how much everyone got to say in the different decisions that were made. In essence it concerns a partition or distribution of the power in the group, what Mulder (1963) calls power distance. If the members of a group have equal say, the power distance is zero. As soon as someone has a greater say than the others (for instance, a trainer vs his pupils) we can say that there is a certain power distance.

Part of the confusion in the discussion about the effects of democratic or autocratic leadership is due to the fact that the distribution of power or influence is confounded with the *degree* of social-emotional or goal-directed leadership, which can vary independently. It is conceivable that in the case of a concentration of all power in one person (very great power distance) a high degree of social-emotional as well as goal-directed leadership is exercised. The greatest pitfall is given in the possible confusion of a high degree of social-emotional leadership (care of the people) and a small power distance. The fact that someone pays much attention to his people and their relationships need not mean that he will allow them to participate in decision making. A kindly disposed patriarch is the classical example of a great power distance at a high degree of social-emotional leadership.

In the present research the only variable will be the degree to which someone participates in the set-up of a training session. To the extent someone participates more in decision making, the power distance to the trainer is smaller. The degree of social-emotional and goal-directed leadership is held constant as much as possible. As mentioned above, the confusion in the discussion about democracy-autocracy is not only a consequence of a mingling of different concepts, but also of the fact which is often lost sight of, that to participate or not to participate can have a different effect on different variables. Hence, in this research the effect of the difference in power distance is investigated for five dependent variables.

Hypothesis I

As the power distance decreases, the less powerful person (P) will gradually demonstrate more *power aspiration*.

This is the original hypothesis of Mulder's power distance reduction theory (1958, 1963). An extensive survey of the literature demonstrated that up to then no connection was made between power distance and

power aspiration with the exception of the studies by Mulder and Veen (Veen, 1970; Part I).

Hypothesis II

As the power distance decreases, the P will gradually demonstrate a great *sympathy* for and *identification* with the more powerful other (A).

This hypothesis was previously demonstrated by Mulder (1963) and Mulder *et al.* (1964, 1966).

Hypothesis III

1. As the power distance decreases, P will gradually be more satisfied and will feel more attracted to his group (AtG).
2. This effect will be manifested foremost and will be strongest in the satisfaction which is narrowly connected to aspects of the first part of the participation procedure.

For this hypothesis, much support is found in the literature. More power gives more satisfaction (Hoffman *et al.*, 1963; Morse & Reimer, 1956; Nealy & Fiedler, 1968; Porter & Lawler, 1965; Strodbeck *et al.*, 1965; Watson, 1965; Watson & Bromberg, 1965). The second part of the hypothesis was formulated on the basis of contradictory results in the experiments by Mulder and co-workers. (See Veen, 1970, Part I.)

Hypothesis IV

As the power distance decreases, P will gradually demonstrate a greater improvement in *performance*.

Analysis of the literature showed that the following conditions have to be met for this hypothesis (Veen, 1970).

The goal directed leadership has to be fairly strong.

The group needs sufficient time to organize itself.

The area in which one is allowed to participate in decision making has to be relevant.

The social-emotional relationships between leader and group have to be fairly good.

Hypothesis V

As soon as feedback about the performance is available, P—within a smaller power distance—will have a more realistic level of aspiration in regard to his own performance, viz. a level of

aspiration which approximates his improvements until that moment.

We assume that P will feel safer at a smaller power distance which enables him to evaluate his possibilities more realistically.

METHOD

Summary

Subjects (Ss) were forty boys, 10–14 years old, who participated in 8 weeks of field hockey training. The training session was held once a week and lasted for 1–1½ hr. The trainers were five boys, 17–18 years old. The research design was a randomized block design, whereby each block consisted of a trainer with a group of Ss. Within each block the Ss were randomly assigned to two conditions: Participation (P) and Non-Participation (NP). The Ss answered questionnaires prior to the training, after the fourth, sixth and eighth training session. Their behavior was observed during each training session.

Manipulation of power distance. The Ss were divided into 5 blocks, roughly homogeneous in age and hockey experience. Each block was divided into 2 groups: Participation (P) and Non-Participation (NP). Both groups were trained by the same trainer. Each training group consisted of 3–5 boys. The P groups had their training session on the day preceding the NP groups. These NP groups executed the same program that was scheduled by their corresponding P group. In this way the tasks of the groups, P and NP, were the same. Before the P group started a training session, the trainer and the boys conferred about 5–10 min about the program. In general, it may be said that this was an unusual procedure for the boys. The usual procedure is for the trainer to have a training program planned without consultation and to execute it with the boys. In effect, the NP procedure mirrors the normal procedure. Such an artifact naturally threatens to interfere with the experimental course. Three steps were taken to allow a successful introduction of the participative decision-making procedure. (Measures, which in our opinion are necessary in every participative decision-making program.)

Ss received the *information* necessary to make decisions about the content of the training. They were given mimeographed copies with a survey of the different aspects of field hockey, exercises, etc. (Veen, 1970, A.I.). This information was *gradually* disclosed and the area in which decision making was permitted was gradually expanded. In the first three training sessions the Ss could decide only about the aspect of training they wanted (for instance, condition training, stopping, etc.). Thereafter they were also allowed to indicate the kinds of exercises they preferred.

Prior to the research the trainers received extensive instructions and exercises for stimulating the Ss to make proposals and handle eventual conflicts. (The trainers, however, were not informed about the hypotheses. They were told that the aim of the study was to investigate the effects of different training methods.)

Measurement of the dependent variables. Three types of measurement were used: (a) questionnaires, (b) observation during the training sessions, and (c) behavioral measures.

Questionnaires. The main questionnaire, answered after the fourth and sixth training sessions, consisted of three parts (all measurements can be found in Veen, 1970, A.2.).

Part 1: This part was aimed at the attitude and the opinion of the Ss in regard to the possibility of becoming an assistant to the trainer (See below: Behavioral measures). Use was made of questions from 60 pt. rating scales, a semantic differential of 14 items about the assistantship, and an open question about its positive and negative aspects.

Part 2: A 20-item semantic differential about the trainer. Ten items were aimed at assessing the appreciation for the trainer and 10 items at assessing his energy and activity as perceived by the Ss.

Part 3: This part consisted of questions with 60-pt rating scales to check the manipulation of the participative decision making procedure and to measure the satisfaction, the power aspiration, and the aspiration to improve the performance. A projective measure required Ss to draw two points in a square (8×8 cm.), to write a T (trainer) and a B (boy) next to them, and to write something that the boy said to the trainer. The distance between the T and B is considered to be an index of the perceived power distance. The remarks are scored for antagonism, positive feelings, etc. toward the trainer (inter-rater reliability: 90%; compare Mulder *et al.*, 1971).

Observations. An observation scheme containing the following categories was developed:

Social-emotional: positive, negative and informal remarks

Power-relations: suggestions, orders

Other: neutral remarks, self-criticism, questions
(Only verbal behavior was recorded).

The observers recorded "closed incidents"; that is, if a person made four remarks with self-criticism in reference to a single incident, this was recorded only once. Each observer was assigned to two training groups from one block, one in P and one in NP (the observers were not informed

about the hypotheses). The reliability was determined by two methods:

(1) an extra observer circulated among all groups. In this way three checks of about 25 min. were obtained for each observer.

(2) after the fourth training session, all observers observed an extra group (which did not participate in the training program). Winer's r_k (1962, page 124) was used as an index of the inter-observer reliability.

Both methods yielded comparable results. The reliability of the social-emotional categories was not acceptable (smaller than .50). The reliability of the other categories was moderate (neutral, self-criticism \pm .60) to high (questions, orders, total number of remarks \pm .90).

Behavioral measures. (Performance and power aspiration): The performance was measured by having Ss run through an obstacle course (Veen, 1970). The Ss were instructed to complete the track as quickly as possible with as few errors as possible. The score is the time needed plus a number of penalty seconds for each error made.

To be able to measure the power aspiration, Ss were given the opportunity to become an assistant to the trainer. To qualify for this position, Ss had to hand in a program for a training period. The number of programs handed in P and NP is used as the index of power aspiration (for a more detailed description of this measure, see below, procedure).

Personality scales. One to three weeks before the training program started, Ss completed a number of self descriptive personality scales:

- (1) The adjective checklist (ACL) by Gough and Heilbrun (1965). The following indices were calculated by a special method which corrects for the total number of adjectives which each person underlined: Self-confidence (Scf), Achievement (Ach), Dominance (Dom), Aggression (Agg), Autonomy (Aut), Nurturance (Nur), Affiliation (Aff), Abasement (Aba), Deference (Def), and the total number of items checked (Nck).
- (2) The individual prominence scale by Shaw (1959).
- (3) The test anxiety questionnaire by Mandler and Sarason (1952). The short high school version was used.
- (4) The fundamental interpersonal relations orientation (FIRO) by Schutz (1958). Those scales were used which measure the degree to which someone is willing to submit to control and to exercise control.

The purpose of the scales was to investigate possible relationships with the dependent variables.

Procedure. In Fig. 1, a summary of the procedure is given.

T1 ^a	T2	T3	T4	T5	T6	T7	T8
Perf. track	Obser- vation	Obser- vation	Obser- vation	Obser- vation	Obser- vation	Obser- vation	Obser- vation
Intro- duction			Intro- duction		Intro- duction		
Question			Question		Question		Question
Obser- vation			Perf. track		Intro. meas. ass't. Perf. track		

^a T1 = training in week *one*, etc.

Training 1 (T1)

(a) The Ss completed "obstacle course" first before any explanation was given of the purpose of the training program and the research connected to it. The result of the Ss was used as the basis for the measurement of subsequent alterations.

(b) After this, an explanation of the training program was given and the researchers and observers were introduced.

(c) The first questionnaire is completed with a double purpose: to make the Ss acquainted with the rating scales and to obtain information for certain control variables.

(d) Subsequently there was a training session of approximately 30 min. with observation.

The participation procedure is not yet introduced. However, the Ss in P were told that next time they would be allowed to make their own program.

T2 and T3

(a) In P each group draws up its own program with the trainer. After that follows a training session of approximately 1 hr in which there are two observation periods of 25 min.

(b) In NP the training session starts immediately. The program previously drawn up by the corresponding P group is executed.

T4

(a) The P groups make their own program. Moreover, they are given the opportunity to indicate the exercises for the various training parts. However, no strong pressure was put on them in this respect.

(b) Training with observation in which the NP group executes the program of the corresponding P group (including eventual exercises).

(c) After the training sessions the Ss were told that from each group one boy could become an assistant to the trainer in three weeks. The negative and positive aspects of the assistantship were clearly outlined (for instance, you have a greater say but also more work to do). As mentioned before, this is one of the most important opportunities to measure power aspiration.

(d) The Ss answered the three-part questionnaire (see above: Questionnaires).

(e) Finally the Ss complete the obstacle course for the second time. By now they know the score of their first performance.

T5

(a) The P groups set up their own program, indicating the training aspects and the exercises.

(b) Training with observation in P and NP.

T6

(a) In P the Ss set up their own program.

(b) Training with observation for P and NP.

(c) The three-part questionnaire is answered (See T4).

(d) After completion of the questionnaire, the Ss were told that the appointment of the assistant was arranged as follows: anyone who wants to become an assistant is to draw up at home a program for one training session and to hand this in. In each group the one who draws up the best program is going to be the assistant.

(e) The Ss complete the obstacle course for the third time. They know the score of T4.

After T6 the main part of the research is finished. T7 and T8 serve to prove the reality of the assistantship. After T8 a last questionnaire was completed with questions about the purpose of the research and suspicions about the existence of another condition.

Design and analysis of the data. As mentioned above, the design was a randomized block design. Each block consists of a trainer and 8-10 boys. The boys are divided at random over the P and NP condition. Ultimately 21 boys remained in P and 19 boys in NP. The main body of the data (questionnaire after T-4 and T-6) is analyzed with a variance analysis with three factors. These are:

Factor A—trainers and group: five levels. This factor is mainly included to extract the error variance connected with the effect of different trainers with their groups.

Factor B—Participation (P,b1) and Non-participation (NP,b2)

Factor C—Time (T4,C1; T6,C2)

This is "case II" of Winer's three-factor experiments with repeated measurements on one factor (Winer, 1962, page 337). The design is complicated by the unequal N-cell. Therefore the unweighted means solution is used (Winer, page 374). In order to test the differences between cells after significant interaction, "simple effects" were computed (Winer, page 174, 340). As this research was mainly concerned with the B and BC effects, the report of the results will exclusively deal with BC tables.

We give a fictitious example in Fig. 2.

A. 33: Do you like your trainer?							
		b1		b2		c	
c1		(1)	30	(4)	40	(7)	35
c2		(2)	40	(5)	30	(8)	35
b		(3)	35	(6)	35	(9)	35
F_B 1				$F_{B(b1)} = 6.00, p .05$			
F_C 1				$F_{B(c2)} = 5.50, p .05$			
$F_{BC} = 14, p .005$				$F_{C(b1)} = 6.30, p .05$			
df of all F-tests: 1, 30				$F_{C(b2)} = 5.80, p .05$			

* F_b gives the value of the F-test on B (cell 3 vs 6).

$F_{B(c1)}$ is the F -test of the difference between b1 and b2 at level c1 (cell 1 vs 4); $F_{c(b1)}$ is the F -test of the difference between c1 and c2 on level b1 (cell 1 vs cell 2); etc.

The analysis of the observations uses a 2×5 analysis of variance with factors B(P-NP) and C(T2-T6).

The A-factor is eliminated because the unit of analysis is a group score, every group being observed only by one observer.

RESULTS

Controls

(a) To be able to ascribe eventual effects to the conditions as such and not to a comparison of the conditions among themselves by the Ss, it was necessary that the Ss had no suspicions about the existence of the other condition. After T8 the question was asked whether the group who had their training session on a different day had an equal influence on the trainer and on which day groups eventually might have a greater influence. Twenty-three out of 29 answers indicate that the groups have an equal influence on different days. Of the 6 Ss who thought that groups

on some days had a greater influence, 4 indicated that this would be an NP group. A question about the purpose of the research did not yield an indication that Ss had suspicions about the existence of the N and NP conditions. From both data, it was concluded with a reasonable degree of probability that the Ss were not aware of the existence of the conditions.

(b) Social-emotional and goal-directed leadership of the trainers: As stated in the introduction, the aim of this research was to vary only the power distance and to keep the social and goal-directed leadership constant in P and NP. In the training of the trainers themselves, much attention was paid to make them aware of the consequences of a certain behavior on the relationships with others. The trainers learned as well how they had to solve possible conflicts. As a control of whether this in fact lead to the desired equality in behavior, the trainers completed two 7-point rating scales after each training session: (1) how friendly is the trainer with regard to the group, and (2) estimate the relationship between trainer and group (good bad). The means of these ratings show almost no difference between P and NP [(1) 4.96-4.64; (2) 4.96-4.48]. Moreover (anticipating the results) it can be said that of the many measurements of the sympathy for the trainer (hypothesis II), almost none yielded a difference between N and NP. The equality of goal-directed leadership is mainly secured by the rigorous equality in the training program of the corresponding P and NP groups. An additional check is available in the ten items about the task activity in the semantic differential about the trainer (part 2 of the questionnaire after T4 and T6). There is neither a significant difference between N and NP on any single item nor in the total score over all ten items.

Conclusion: the measurements do not justify the supposition of any difference between P and NP in regard to social-emotional and goal-directed behavior.

(c) Objective realization of the participation procedure in P: Two things had to be achieved in order to insure the success of the participation procedure: (1) every group had to make enough proposals to be able to compose a program, (2) the number of proposals per subject within a group should not be too different in order to give each participant the impression of having a real say in the program. This could be checked by the record which the trainers kept of the proposals. The mean number of proposals per person in the different training groups was: $a_1 = 2.2$; $a_2 = 2.8$; $a_3 = 4.3$; $a_4 = 2.2$; $a_5 = 2.0$. (n is respectively 4, 4, 3, 5, 5). Therefore the groups made 9-13 proposals during each training session of which 6-8 could be executed. The difference in number of proposals between the different subjects within the group is apparent from the following data:

Only five out of 21 Ss in P had an average number of proposals per training session slightly lower than 2 (lowest 1.4).

The average number of proposals per subject per training session was 2.6. From these data we may conclude that the participation in the design of the training program is successful.

(d) In addition to the objective realization of participation, it is desirable that the difference between P and NP is perceived by the Ss (compare Lowin, 1968). Questions 9 and 10 of the 3-part questionnaire are directly aimed at this point.

Question 9: do you have the feeling that your trainer in setting up today's training program took your opinion into account? (very much so just a little).

Question 10: how much did you contribute to today's program of your group? (very much . . . just a little).

Both questions demonstrate a significant difference between P and NP (means: Question 9: $P = 49.6$, $NP = 33.2$, $p < .001$. Question 10: $P = 37.7$, $NP = 19.9$, $p < .001$). Another question (does your word carry any weight with the other boys in the group? very much . . . absolutely nothing) reveals that the participation procedure is felt as a way of influencing the other boys (means: $P = 29.5$, $NP = 22.8$, $F_{B(c1)} < 1$, $F_{B(c2)} = 12.54$, $p < .005$). So the difference P-NP is only significant after T6. The projective measurement, requesting the boys to put a T (trainer) and B (boys) in a square provides a last check. The distance in mm between T and B is an index of the perceived distance. This is significantly smaller in P than in NP (means: $P = 38.9$, $NP = 54.4$; $F_B = 7.58$, p is smaller than .025). All these results justify the conclusion that the subjective experience of the Ss is in agreement with the objectively registered difference between the conditions N and NP.

Power aspiration. Hypothesis I stated that the Ss in P (a smaller power distance) will gradually demonstrate more power aspiration. Question 1 (after T4 and T6) was: do you want to be an assistant to your trainer? (very much . . . little desire). Table 1 gives the results for this question.

There appears to be a strong BC interaction. Initially (after T4 = C1) the aspiration in NP is somewhat higher than in P. The difference is statistically significant. (F_{Bc1}). After T6 the reverse is true. The aspiration in P is now much higher than in NP (F_{Bc2}). This change has to be ascribed mainly to a drop in aspiration in NP (F_{Cb2}). These data confirm hypothesis I, although the higher aspiration in NP after T4 raises some questions.

As mentioned above, the Ss were asked to make out and hand in programs if they wanted to become an assistant. The number of programs

TABLE 1

Q.1: Do you want to be an assistant to your trainer?			
	b1(P)	b2(NP)	C
C1(T4)	35.1	44.3	39.5
C2(T6)	43.6	26.0	35.3
B	39.4	35.2	37.4
B: $F = <1, p$; B(c1): $F = 4.66, p < .05$ C: $F = 4.60, p < .05$; B(c2): $F = 11.19, p < .005$ BC: $F = 24.55, p < .001$; C(b1): $F = 3.99, p < .10$ df F -tests = 1, 30; C(b2): $F = 25.2, p < .001$			

handed in is considered an index of power aspiration. The number of programs handed in was disappointingly low. The distribution over P (6) and NP (1) supports hypothesis I (Fisher-test $P = .0465$, one-sided). Three other questions, although not directly connected with the assistantship, were aimed at power aspiration: question 20—would you like to have more influence on your trainer?; question 19—would you like to have a greater say in your group?; question 21—would you like to be a captain of a team? All these questions manifest a significant difference between P and NP that is greatest after T6, which therefore confirms hypothesis I (See Tables 2, 3 and 4).

In the projective question #23, the boys were asked to write down a remark made by a boy to a trainer. The answers were scored by a scheme of categories developed in previous research (Mulder, Veen *et al.*, 1971; interscorer agreement 89%). The category "exercise of power" in this

TABLE 2

Q.20: Would you like to have influence on your trainer?			
	b1(P)	b2(NP)	C
C1(T4)	27.1	28.3	27.7
C2(T6)	33.7	20.5	27.4
B	30.4	24.4	27.6
B: $F = 3.11, p < .10$; B(c1): $F = <1, p$ C: $F = <1, p$; B(c2): $F = 8.80, p < .01$ BC: $F = 6.19, p < .025$; C(b1): $F = 1.73, p$ df F -tests: 1, 30; C(b2): $F = 4.86, p < .05$			

TABLE 3

Q.19: Would you like to have influence on the boys?			
	b1(P)	b2(NP)	C
C1(T4)	28.6	21.9	25.4
C2(T6)	28.7	15.7	22.5
B	28.6	18.8	24.0
B: $F = 8.99, p < .01$; B(c1): $F = 2.16, p$ C: $F = 1.26, p$; B(c2): $F = 8.13, p < .01$ BC: $F = <1, p$; C(b1): $F = <1, p$ df F -tests: 1, 30; C(b2): $F = 2.09, p$			

scheme is relevant for hypothesis I. Although the frequencies are low the results are of great importance because the same phenomenon is measured with another type of measurement. The frequencies for this category appear in Table 5.

Again we find support for hypothesis I, because of the fact that 10 remarks about exercise of power were made in P and 3 in NP. As was the case in all the other measurements, the greatest difference appears after T6.

All in all we conclude that hypothesis I is strongly supported by the facts. However, each individual measurement can be criticized, but, as Webb *et al.* (1966, page 3) put it: "If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it." It seems appropriate however to call attention to the fact that only the final effect (T6) can be predicted with some certainty. The preceding patterns vary considerably.

TABLE 4

Q.21: Would you like to be captain of a team?			
	b1(P)	b2(NP)	C
C1(T4)	37.5	29.6	33.8
C2(T6)	36.2	21.6	29.3
B	36.9	25.6	31.5
B: $F = 4.40, p < .05$; B(c1): $F = 2.02, p$ C: $F = 5.32, p < .05$; B(c2): $F = 5.80, p .025$ BC: $F = 1.46, p$; C(b1): $F = <1, p$ df F -tests: 1, 30; C(b2): $F = 6.18, p .025$			

TABLE 5
Q.23 Frequencies
"to exercise power over"

	b1(P)	b2(NP)	
c1(T4)	3	2	5
c2(T6)	7	1	8
	10	3	13

Identification and sympathy. Hypothesis II predicted that in a smaller power distance (P) the Ss would develop a greater sympathy for and identification with the trainer. These variables were measured by ten evaluation items from the semantic differential about the trainer. A weighted score was computed from them. (See Veen, 1970, page 172 and following.)

Question 17: do you like the trainer?

The category "positive feelings" of the projective question #23 (what remarks does the boy make to the trainer).

The observation categories (positive, negative, and informal). Only the results for the category "positive feelings" gives some support to hypothesis II (P = 10 remarks; NP = 4). All other indices do not support the hypothesis. Therefore we conclude that hypothesis II is not confirmed.

Satisfaction and attraction to the group (AtG). Hypothesis III predicted that the Ss in P would gradually show more satisfaction in contradistinction to NP. This effect would be strongest for the components of satisfaction most closely connected to aspects of the participation procedure.

Three questions were aimed at the general satisfaction and AtG:

question 11 How do you like attending the training session?

question 12 How do you like the training in general?

(Like it very much annoying)

question 15 How do you like the boys in your group?

(Can't be better very annoying)

These questions do not yield any difference between P and NP.

Two questions (13, 14) were aimed at the satisfaction connected to aspects of the participation procedure. Both questions indicate more satisfaction in P after T6 than in NP (F_{BC2}). Therefore this part of hypothesis III is confirmed (See Tables 6 and 7).

The number of remarks in the category "antagonism" of question 23—

TABLE 6

Q.13: May the training be improved?			
	b1(P)	b2(NP)	C
C1(T4)	34.9	36.0	35.4
C2(T6)	36.8	25.2	31.3
B	35.8	30.6	33.3
B: $F = 1.47, p$; $B(c1): F = <1, p$ C: $F = 2.39, p$; $B(c2): F = 4.67, p < .05$ BC: $F = 3.45, p < .10$; $C(b1): F = <1, p$ df F -tests: 1, 30; $C(b2): F = 5.79, p < .025$			

which mainly expresses dissatisfaction with the content of the training—support hypothesis III ($P = 6, NP = 20$). Also in this case the result is the greatest after T6 ($P = 3, NP = 12$).

We conclude that hypothesis III is not confirmed in regard to the general satisfaction and AtG. However, the hypothesis is confirmed as far as the aspects are concerned which are more directly connected to the participation procedure or its consequences (program).

TABLE 7

Q.14: Are you satisfied with the training-program?			
	b1(P)	b2(NP)	C
C1(T4)	42.5	40.2	41.4
C2(T6)	43.6	29.2	36.7
B	43.0	34.7	39.0
B: $F = 5.98, p < .025$; $B(c1): F = <1, p$ C: $F = 3.17, p < .10$; $B(c2): F = 9.55, p < .005$ BC: $F = 3.78, p < .10$; $C(b1): F = <1, p$ df F -tests: 1, 30; $C(b2): F = 6.94, p < .025$			

Performances

Hypothesis IV predicts that a greater performance improvement will gradually be shown in P. Performance improvement was measured by the score difference between the results in the obstacle course after T4 and T6 and the result after T1. Table 8 shows the results of P and NP. The difference is significant in the predicted direction. P (b1) manifests

TABLE 8

Improvement of Performance from T1 on (in sec.)*			
	B1(P)	b2(NP)	C
C1(T4)	4.95	1.63	3.38
C2(T6)	4.86	-1.05	2.05
B	4.90	.29	2.71
B: $F = 4.60, p < .05$; B(c1): $F = 2.43, p$ C: $F = 3.23, p < .10$; B(c2): $F = 6.29, p < .025$ BC: $F = 2.22, p$; C(b1): $F = 1, p$ df F -tests: 1, 30; C(b2): $F = 5.41, p < .05$			

* The figures in this table give the differences in performance between T0-T4(c1) and T0-T6(c2).

a greater performance improvement than NP(b2). Again the difference is most pronounced after T6 (B[c2]-effect).

It is possible that by an unfortunate artifact in the randomization procedure, the Ss in P had a greater natural ability to play field hockey. If this were the case, an even greater difference between P and NP should be apparent after one year. To check this, Ss completed the obstacle course once again after one year. We see that the results are practically the same in P and NP (respectively 17.9 and 18.5 seconds). The improvement since T6 is: P = 2.5 seconds and NP = 5.5 seconds. It is obvious, therefore, that the Ss in NP recovered during this year from the inferior performance they had clearly shown during the training program. Thus it seems reasonable in every respect to ascribe the differences that developed in performance improvement to the participation procedure.

Aspiration Level in Regard to Performances

Hypothesis V predicts that the Ss in P will have a more realistic level of aspiration when feedback about the real improvement is available. Question 22 was: If your performance last time was — seconds, what time do you intend to make this time? (— seconds to be filled in). After T4 the Ss were told the result of their performance at T0; after T6, the results of T4. Therefore they could observe their improvement from T0-T4 before completing the obstacle course after T6. To test hypothesis V we need the absolute difference between the level of aspiration before T6 and the objective improvement from T0-T4. According to the hypothesis the difference should be smaller in P than in NP. The results confirm hypothesis V. The means for P = 3.80, for NP = 7.05 ($F = 7.38, p < .025$).

Personality Scales

There is no systematic relationship between the personality scales and any of the other variables. A number of reasons may account for this. The most plausible seem to be: (a) the self reporting scales are on too high a level of verbalization for this age group; (b) the validity of the American scales is low for the Dutch population.

DISCUSSION

Internal Validity

The first problem in any research is naturally the internal validity: to what degree can the effects be ascribed to the experimental variables? By the choice of a "real experimental design" (Campbell and Stanley, 1963, page 175) a number of interpretations can be excluded like history, maturation, testing, instrumentation, and mortality. Based on the checks, one may state with a reasonable amount of confidence that the social-emotional and goal-directed behavior of the trainers was equal in both conditions. The participation procedure was objectively and subjectively realized which, according to Lowin (1968), is a necessary condition for any research on participation. All in all, the conclusion can be drawn that the internal validity is high enough to ascribe the effects to the P-NP variable.

RESULTS

The most striking feature in the results is undoubtedly the declining trend of a number of variables in NP. A possible cause of this is found in the high level of expectation of the Ss in regard to the training, manifested in a number of questions that were asked in advance of the first training session. On an average, scores of 45 were reached on these questions out of a maximum possible score of 60. The questions on "satisfaction" (questions 11-14), which were asked after T4 and T6, show an interesting correlation pattern with the "expectation questions." After T4, in P and NP, there is a significant correlation between expectations about the training and the general satisfaction at that moment (T4). Only in NP there appears to be correlation between the expectations and the satisfaction specifically connected to the participation procedure. This pattern was drastically altered after T6. In P the expectations correlate very highly with the specific satisfaction (.80 and .70) and to a lesser degree with the general satisfaction (.60). In NP there is no correlation between the initial expectations and the satisfaction. We interpret this phenomenon as follows: in P the high expectations are essentially confirmed. In NP, however, the expectations are confirmed

initially but somewhere between T4 and T6, disappointment occurs and the experiences at T6 are no longer related to the original expectation. Although this is a plausible explanation, and the only one which we are able to support with empirical relationships, it does not seem completely satisfying. It seems particularly remarkable that the results in NP, where traditional methods of training were used, shows a slight decline. One explanation is that the obstacle course measures mainly "motivation" and only to a lesser degree "ability," two components which appear in every performance and which reflect the influence of desire and ability. It is difficult to say whether this explanation is valid here. The dominance of the motivational component is in some degree acceptable because of the fact that many boys had a very good performance on their first trial.

External Validity

The ultimate value of any research is determined by the extent to which the results can be generalized (external validity). In judging this matter, we distinguish three aspects: (a) the strength of the manipulated variable (b) the naturalness of the situation (c) the level of abstraction of the manipulation.

Ad (a) The strength of the manipulation is probably weak. The total participation procedure takes only 5-10 min. per training session. On the other hand, the topics in which Ss participate in decision making are of vital importance for themselves, which is one of the most important conditions for participative procedures (French *et al.*, 1960; Lowin, 1968). The fact that the effects become visible through all the noise of the field justifies the expectation that the manipulation was strong enough.

Ad (b) A problem in almost any research in the social sciences is the fact that the research changes something in the situation under investigation. There is the danger that the Ss start behaving differently. A number of factors ensures that this is unlikely in this situation:

(1) Playing field hockey is much more important for these Ss than the research. It was striking for instance, that nobody talked about the research in the dressing room and immediately after the training. The coming game and the team formation were much "hotter" topics of conversation.

(2) The longitudinal character of the research allows the Ss to get accustomed to the research and the researchers.

(3) The Ss were at an early age which entails a certain naivete. This is the last factor which contributes to the fact that the Ss

remain themselves. (Compare Weick, 1968, pages 372 and following.) On the other hand, it remains a disadvantage that obstructive questionnaires had to be used.

- Ad (c) The level of abstraction of the manipulated variable and of the situation is very low. A series of concrete factors make up the background of this investigation: the situation has a spare-time character and the Ss come voluntarily and for their own pleasure. The age is very young: 10-14 years old. The distance between boys and trainer is relatively small. This means that the results can, in principle, be generalized only to situations with equal characteristics. Only if the results of this research can be correlated with the results of other studies will it be possible to say more about the general validity of the results. This, however, is outside the scope of this article. An indication is given in the literature cited in drawing up the hypotheses. The conclusion which we draw from A, B and C is: the results can be ascribed with confidence to the experimental variable (P-NP). Moreover, it seems probable that the Ss behaved naturally. Generalization, however, is possible only to comparable situations.

REFERENCES

- BALES, R. F. The equilibrium problem in small groups. In Parsons, T., Bales, R. F., and Shils, E. A. (Eds.), *Working Papers in the Theory of Action*, Glencoe, IL: Free Press, 1953, 111-161.
- CAMPBELL, D. T., & STANLEY, J. C. Experimental and quasi-experimental designs for research on teaching. In Gage, N. L. (Ed.), *Handbook of Research on Teaching*, Chicago: Rand McNally, 1963, 171-246.
- CARTER, L. F. Leadership and small group behavior. In Sheriff, M., and Wilson, M. O. (Eds.), *Group Relations at the Crossroads*, New York: Harper, 1953, 257-284.
- COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. In *Human Relations*, 1948, 1, 512-532.
- FRENCH, J. R. P., JR., ISRAEL, J., & AS, D. An experiment on participation in a Norwegian factory. In *Human Relations*, 1960, 13, 3-19.
- GOUGH, H. G., & HEILBRUN, A. B. *The Adjective Checklist: Manual*, Palo Alto, CA: Consulting Psychologists Press, 1965.
- HOFFMAN, R. L., BURKE, R. J., & MAIER, N. R. F. Participation, influence and satisfaction among members of problem solving groups. In *Psychological Report*, 1965, 16, 661-667.
- KAHN, R. L., & KATZ, D. Leadership practices in relation to productivity and morale. In Cartwright, D., and Zander, A. (Eds.), *Group Dynamics*, 2nd ed. London: Tavistock, 1960, 554-570.
- LOWIN, A. Participative decision making: A model, literature, critique, and prescriptions for research. In *Organizational Behavior and Human Performance*, 1968, 3, 68-106.
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. In *Journal of Abnormal Social Psychology*, 1952, 47, 166-173.

- MORSE, N. C., & REIMER, E. The experimental change of a major organizational variable. In *Journal of Abnormal and Social Psychology*, 1956, **52**, 120-129.
- MULDER, M. Groepsstructuur, motivatie en prestatie. In *Den Haag*, 1958.
- MULDER, M. Machtsmotief, positieve en negatieve identificatie. In Mulder, M. (Ed.), *Mensen, Groepen, Organisaties*, deel I, Assen: Van Gorcum, 1963, 205-232.
- MULDER, M., VAN DIJK, R., SOUTENDIJK, S., STELWAGEN, T., & VERHAGEN, J. Non-instrumental liking tendencies toward powerful group members. In *Acta Psychologica*, 1964, **22**, 367-386.
- MULDER, ET AL. Structuur en dynamiek van een grote organisatie: een veld-studie op zee. Utrecht: Instituut voor Sociale Psychologie, 1967 Technisch rapport.
- MULDER, M., VEEN, P., HARTSUIKER, D., & WESTERDUIN, T. Cognitive processes in power equalization. In *European Journal of Social Psychology*, 1971, **1**, 107-130.
- NEALY, S. M., & FIEDLER, F. E. Leadership functions of middle managers. In *Psychological Bulletin*, 1968, **70**, 313-329.
- PORTER, L. W., & LAWLER, E. E., III. Properties of organization structure in relation to job attitudes and job behavior. In *Psychological Bulletin*, 1965, **64**, 23-51.
- SCHUTZ, W. C. (FIRO) *A Three-dimensional Theory of Interpersonal Behavior*, New York: Rinehart & Company, 1958.
- SHAW, M. Some effects of individually prominent behavior upon group effectiveness and member satisfaction. In *Journal of Abnormal Social Psychology*, 1959, **59**, 382-386.
- STRODTBECK, F. L., SIMON, R. J., & HAWKINS, C. Social status in jury deliberations. In Steiner, I. D., and Fishbein, M. (Eds.), *Current Studies in Social Psychology*, New York: Holt, Rinehart & Winston, 1965, 333-342.
- VEEN, P. *Meebeshissen, Een Veldexperiment in Een Hockeyclub*, Assen: Van Gorcum, 1970.
- WATSON, D. L. Effects of certain social power structures on communication in task-oriented groups. In *Sociometry*, 1965, **28**, 322-336.
- WATSON, D. L., & BROMBERG, B. Power, communication and position satisfaction in task-oriented groups. In *Journal of Personality and Social Psychology*, 1965, **2**, 859-864.
- WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D., & SECHREST, L. *Unobtrusive measures: Nonreactive Research in the Social Sciences*, Chicago: Rand McNally, 1966.
- WEICK, K. E. Systematic observational methods. In Lindzey, G., and Aronson, E. (Eds.), *Handbook of Social Psychology*, 2nd ed. Reading, MA: Addison-Wesley, 1968, **2**, 357-451.
- WINER, B. J. *Statistical Principles in Experimental Design*, New York: McGraw-Hill, 1962.

RECEIVED: April 29, 1970