# EVALUATION OF BULK HYBRID TESTS FOR PREDICTING PERFORMANCE OF PURE LINE SELECTIONS IN WINTER WHEAT

by

WAYNE LOVELLE FOWLER

B. S., Kansas State College of Agriculture and Applied Science, 1951

---

A THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Agronomy

KANSAS STATE COLLEGE
OF AGRICULTURE AND APPLIED SCIENCE

1953

TABLE OF CONTENTS

# INTRODUCTION

Hybridization, followed by selection of desirable segregates, has come to be the principal method of improving the small grains. Plant breeders have turned from a system involving isolation of the better strains occurring in a variety to a method in which they attempt to combine the desirable characteristics of two strains, varieties, or species.

In a hybridization program workers are faced with the problem of deciding which are the best crosses to make and then choosing the best segregates from the successful crosses. Limitations of funds, land, labor, and time make it impossible for plant breeders to carry out the crossing programs they might like. The number of crosses must generally be curtailed in accordance with the time and resources that can be spent searching for the best segregates among the hybrids. The accepted thing is to carefully choose the parents that are to be used; this often leaves the breeder with the thought that some other cross might have been better.

It is believed to be desirable for a plant breeder to observe and work with as many crosses as possible. Any system that will increase the efficiency with which segregating generations are handled, thereby making it possible to work with more crosses, is worthy of consideration as part of a plant breeding program. Such a system should either make it possible to eliminate entire crosses within a few generations or else reduce the number of plants or lines continued from each cross; this must be accomplished while the probability of discarding potentially valuable segregates remains low.

It is the purpose of this paper to report the results of a study in which the bulk population of each of 45 winter wheat crosses was tested in F3, F4, and F5 generations under the hypothesis that such a test might reliably predict the value of segregates selected in the fifth generation of that bulked cross.

## REVIEW OF LITERATURE

Hayes and Immer (12) and Love (15) describe the handling of hybrid populations by the pedigree, or line, method. This involves selection of individual F2 plants with continued selection of lines during succeeding generations, the progeny of each selection being maintained separately. Nilsson-Ehle devised an alternative, the bulk method, which is discussed by Ackerman and MacKey (1), Florell (7), Hayes and Immer (12), and Love (15). This method, wherein all the F2 seed of a cross is bulked and this bulk population is carried for four or five generations, offers considerable reduction in the effort necessary to maintain a cross. It has been adopted by some workers because it entails less record keeping and handling during early generations than the pedigree system and therefore allows time to work with more crosses. Under the bulk system natural selection is supposed to operate to eliminate low yielding and unsuitable types, increasing the chance of selecting a desirable type after four or five years of bulk propagation.

The bulk method makes it impossible to carry on a genetic study in conjunction with the breeding program because progenies of individual plants and families are not separated. Suneson and Wiebe (2) present evidence from barley and wheat trials in which the yield of a variety was not necessarily a measure of its ability to survive in mixed stands, high yielding varieties being unable to compete as well in a mixture as certain low yielding varieties. They believe that this situation suggests a decided limitation to the bulk method, implying the assumption that the ability to survive in a population is not necessarily a criterion of the yielding capacity of a hybrid. The belief that cold, disease, and other adverse conditions will operate to eliminate types is in agreement with the thought of these writers but they conclude that, in the absence of these adverse conditions, valuable material is likely to be lost from a bulked hybrid population because of competition.

Studies of early generation testing in various crops have not all led to similar conclusions. In the case of corn, a cross-pollinated crop, the value of early testing for yield seems to be well established and of considerable value. Sprague (19) emphasizes the desirability of determining the potential value of an inbred line during the first and/or second selfed generations and describes the method used for making such determinations. Each plant selected for inbreeding is selfed and outcrossed to a suitable tester parent during the first year. A yield trial of the test cross progeny during the succeeding year is used to estimate the relative combining ability of the S0 plants. Undesirable selfed plants are eliminated without further inbreeding, allowing more time and effort to be spent on remaining inbreds, which are known to be from the high yielding portion of the population. The same testing process may be repeated in the S4 generation, final elimination of inbred lines being made before any single crosses are attempted.

Harlan, Martini, and Stevens (9) tested seven generations of 379 bulked barley crosses in unreplicated single-row plots. After seven years the crosses were divided into five yield groups. Selection of individual F8 plants was made in proportion to previous performance, high yielding crosses contributing more selections than low yielding ones. When the selections were grown in a comparative yield test the following year it was found that those from the low yielding group were consistently low yielding, those from the high yielding group continued to yield high, and that the relationship between yield groups was maintained by the selections from them. It was concluded that the preselection classification had been effective and that the low yielding group of crosses, constituted of the poorer types, could just as well have been discarded prior to selection.

Immer (13) believed that determination of the average yield of barley crosses made on bulk F2 and F3 generation populations would be of value in detecting the better crosses of a group. Only the high yielding crosses, presumably having the greater proportion of high yielding genotypes, would be retained for selection. It is pointed out that a yield test in the F1 generation would not be suitable because the small amount of seed available from each cross compels space planting. It is known that the yield of a cross or variety in space planted rows is unlike the yield performance of the same cross or variety in drilled rows. An indication of the questionable value of yield results secured during the F2 generation is given by Grafius, Nelson, and Dirks (8) who found most of the variance in bulked F2 barley progenies to be nonheritable. The heritability of the more homozygous F3 generation bulks was greater, indicating that selection for yield during the F3 generation might be of some value.

Harrington (11), assuming that the degree of heterosis shown by early generations of wheat crosses would be a true indication of the yield value of a cross, tested six bulk populations in the F2 and F3 generations in replicated yield trials and compared the results with the yield of selected lines in the F6, F7, and F8 generations. He found that the test of the F2 bulk generation accurately evaluated the yielding ability of each cross as shown by its segregates; the test of the F3 bulk generation was said to have supplementary value in supporting the conclusions drawn from results in the F2 generation. It was anticipated that F2 bulk hybrid tests would be of less value in studying milling and baking quality, disease resistance, and reaction to weather conditions.

Harrington (10) reported, in an earlier paper, a somewhat different approach in which a random sample of F2 plants of a cross was studied. When only six lines out of 40,000 F2 plants of a Marquillo x Marquis cross remained after five years of selection the remnant seed of a random sample of F2 plants was sown. An analysis of these plants, which could not include baking tests, indicated that about seven good lines could have been expected from the original group of F2 plants. The analysis of the F2 generation had correctly predicted stem rust reaction, earliness, plant height, and seed characters of the lines but had been misleading as to yield, probably because the yield of an individual plant tells little about the inherited yielding ability of that plant. It was pointed out that such an analysis of the F2 generation of each cross would be costly and that such a system did not facilitate direct comparisons between crosses.

Bjaanes (3) reported on an F2 progeny test of four spring wheat crosses made in Norway. The progenies, called "F2-families," of 28 selected F2 plants were tested in yield trials for four years. When selections made in the F6 generation were tested it was found that those from low yielding F2-families were, in the majority of cases, low yielding and that the one F2-family found to be superior in preliminary tests had given, in the majority, high yielding lines. This was considered fair proof that yield trials during segregating generations furnish a good basis for choosing F2-families worthy of being used for line selection.

Atkins and Murphy (2) classified ten oat crosses as high or low yielding on the basis of replicated rod row tests of bulk F2 through F6 generations. Correlation of yield from generation to generation was low. Fifty selections were made from each cross in the F7 generation. When the results of a cubic lattice yield trial were analyzed the following year it was found that the preselection classification had not been successful; as many high yielding segregates came from the crosses classified as low yielding as from the high yielding crosses. In fact, the two crosses that produced the greatest number of high yielding segregates had been classified as low and would have been discarded in a program based on early testing. The authors did find test weight to be reliably predicted in the same study and concluded that the test weight of one or two early generations would be useful in determining the value of a cross for this one characteristic.

Kalton (14) evaluated the bulk populations of F2, F3, and F4 generations of 25 soybean crosses for yield, maturity, plant height, and lodging. Yield differences between crosses were inconsistent from generation to generation; therefore, detection and elimination of low yielding crosses on the basis of one test would not have been safe. Maturity, plant height, and lodging was relatively constant from generation to generation, however, and it seems that one test would be sufficient to evaluate these characters for the early generations of a cross. Plant height and maturity measurements on spaced F2 plants provided a good estimate of the F3 and F4 progeny performance for these characters but yield measurements made on the same plants were of no value in predicting the yield of F3 and F4 progenies.

Mahmud and Kramer (16) reached a similar conclusion in a less extensive test on soybeans. The effect of environment on the early tests was believed to be so great as to reduce heritability estimates for yield to negligible values while those for maturity and plant height remained high.

Weiss (22) and Weiss, Weber, and Kalton (23) found no relation between the degree of heterosis shown by the F1 plants of any of 17 soybean crosses and yield of bulks in any succeeding generation from F2 through F5; nor was the degree of heterosis of F1 plants related to the yield of selections made in the F5 generation. Date of maturity, height, and lodging evaluations of the F1 plants were found to be of no value for prediction. Height and maturity measurements on spaced F2 plants of a cross were fairly effective in predicting the performance of subsequent selections from that cross; lodging could not be evaluated because spaced plants and those in drilled rows do not react the same. Tests of bulk populations of F2 through F5 generations were not reliable in predicting yield or maturity date of selections from crosses but lodging and height evaluations were consistently indicative of the performance of subsequent selections.

Ackerman and MacKey (1) report that the present procedure of breeding self-fertilized plants at Svalof, Sweden, is to make selections during the F2 or F3 generation, test these lines for a few years, and make final selection from the lines shown to be best in those early trials. They have shown that the variation between lines may be significant while that within lines is low enough that all the components of a line can be judged by the average of that line. The Swedish studies involve such characteristics as maturity, strength of straw, and disease resistance but evidently not yield because the writers comment on American studies involving yield predictions based on early generation testing. They contend that the factors influencing yield are non-additive and that the effect of each factor decreases with increasing numbers of factors, making it questionable whether early generation testing for yield can be of value. Under Swedish conditions, they say, such testing might isolate the very good and the very poor crosses but that year to year climatic changes, coupled with low fertility of certain hybrids and shortage of seed with which to conduct yield trials during early generations, would frequently distort the yield results.

The above reviews show the diverse results that have been reported on early generation tests of various self-fertilized crops. It is not possible to conclude what, if any, characters can be predicted, nor is it clear how and in what generations one should proceed to evaluate a character if it were found to be of value for predictive purposes. The situation is quite unlike that in corn where early testing is well established. Further studies are evidently needed before the value of early generation testing can be accurately determined.

## MATERIALS AND METHODS

### Parents and Crosses Studied

Ten winter wheat varieties were crossed in all possible ways during the 1942 greenhouse and field season at Manhattan, Kansas. Table 1 lists the parental varieties and indicates the diverse types included. Disease and insect reaction of the parents may be summarized as follows:

Loose smut (*Ustilago tritici*) -- Pawnee highly resistant, all others susceptible.

Bunt (*Tilletia foetida*) -- Comanche and Nebred highly resistant, Pawnee and Marquillo-Oro moderately resistant, others susceptible.

Leaf rust (*Puccinia triticina*) -- all susceptible as adult plants under Kansas field conditions.

Stem rust (*Puccinia graminis tritici*) -- all susceptible; Blackhull, Chiefkan, and RedChief show tolerance.

Hessian fly (*Phytophaga destructor*) -- Pawnee and Marquillo-Oro show resistance or tolerance, others susceptible.

All the parents may be considered representative of varieties that are, or have been, commercially important in the wheat belt of Kansas and adjoining states. Early Blackhull-Tenmarq is a sister selection of the variety Wichita. Marquillo-Oro is the only variety of the ten parents that was never released to wheat growers. It is not equal to other available varieties in some agronomic characteristics but has been important in the Kansas wheat breeding program because of its hessian fly resistance.

Reciprocal crosses were found not to differ; therefore, 45 crosses were included in all later studies. At the time of crossing it was planned to bulk the F2 seed of each cross and conduct yield trials of bulk F2, F3, and F4 generation crosses. Such a study was expected to indicate whether or not the tests of bulk populations would predict the value of a cross as plant breeding material and to give information on the combining ability of the ten varieties involved.

## Characteristics Studied

F1 plants of each cross were grown both in the greenhouse and in the field during the 1943 season. All the F2 seed harvested from the F1 plants of each cross was bulked and used to plant a variable number of thinly-sown four-foot rows in the fall of 1943. Notes taken on the F2 generation bulks the following spring and summer included first head, test weight, and yield.

The F3 generation was the first to be entered in comparative yield trials. In 1946 and 1947 the 45 F3 generation bulks, along with the ten parents, were grown in randomized single eight-foot rows replicated ten times. In 1948 the same bulks and parents were grown in five replications of randomized two-row plots eight

Table 1. Characteristics of the 10 parental winter wheat varieties under Manhattan, Kansas, conditions.

| Parent | Identification No. | Yield | Plant Height | Maturity | Test Weight |
|---|---|---|---|---|---|
| Blackhull | CI 6251 | med. | tall | late | med. |
| Cheyenne | CI 8885 | low | short | late | low |
| Chiefkan | CI 11754 | med. | tall | late | high |
| Comanche | CI 11673 | med. | med. | med.-early | med. |
| Early Blackhull-Tenmarq | Ks 2757 | high | short | early | high |
| Marquillo-Oro | CI 11979 | med. | tall | late | low |
| Nebred | CI 10094 | low | short | late | low |
| Pawnee | CI 11669 | high | short | med.-early | med. |
| RedChief | CI 12109 | med. | tall | med. | high |
| Tenmarq | CI 6936 | med. | med. | med. | low |

feet long. The seed for planting the 1946, 1947, and 1948 F3 generation bulk trials came from the same source of 1944 seed, which was kept in cold storage to maintain its viability. Table 2 lists the 45 crosses and ten parents under the entry number assigned to each in all F3 and subsequent generation bulk tests. Observations recorded for the F3 generation bulk nurseries include first head (or date of flowering), test weight, plant height, and yield.

Comparative yield tests of F4 generation bulks and parents were made in 1947 and 1948, the seed coming from the 1946 F3 nursery in both cases. The 1947 nursery was grown as a randomized complete block with five replications of single eight-foot rows; the same design was used in 1948, but consisted of two-row eight-foot plots replicated five times.

Table 2. Entry number of, and number of selections tested from, 10 varieties of winter wheat and 45 crosses among them grown at Manhattan, Kansas, 1946-52.

| Entry No., 1946-48 Tests | Kind | No. Selections Tested, 1951 and 1952 |
|---|---|---|
| 1 | Tenmarq, CI 6936 | 2 |
| 2 | Tenmarq x Comanche | 7 |
| 3 | Tenmarq x Cheyenne | 7 |
| 4 | Tenmarq x Chiefkan | 7 |
| 5 | Tenmarq x RedChief | 7 |
| 6 | Tenmarq x Pawnee | 7 |
| 7 | Tenmarq x Blackhull | 7 |
| 8 | Tenmarq x Nebred | 8 |
| 9 | Tenmarq x Marquillo-Oro | 7 |
| 10 | Tenmarq x Early Blackhull-Tenmarq | 7 |
| 11 | Comanche, CI 11673 | 2 |
| 12 | Comanche x Cheyenne | 7 |
| 13 | Comanche x Chiefkan | 8 |
| 14 | Comanche x RedChief | 7 |
| 15 | Comanche x Pawnee | 7 |
| 16 | Comanche x Blackhull | 8 |
| 17 | Comanche x Nebred | 7 |
| 18 | Comanche x Marquillo-Oro | 7 |
| 19 | Comanche x Early Blackhull-Tenmarq | 7 |
| 20 | Cheyenne, CI 8885 | 2 |
| 21 | Cheyenne x Chiefkan | 7 |
| 22 | Cheyenne x RedChief | 8 |
| 23 | Cheyenne x Pawnee | 7 |
| 24 | Cheyenne x Blackhull | 7 |
| 25 | Cheyenne x Nebred | 7 |
| 26 | Cheyenne x Marquillo-Oro | 7 |
| 27 | Cheyenne x Early Blackhull-Tenmarq | 7 |
| 28 | Chiefkan, CI 11754 | 2 |
| 29 | Chiefkan x RedChief | 7 |
| 30 | Chiefkan x Pawnee | 8 |
| 31 | Chiefkan x Blackhull | 7 |
| 32 | Chiefkan x Nebred | 7 |
| 33 | Chiefkan x Marquillo-Oro | 7 |
| 34 | Chiefkan x Early Blackhull-Tenmarq | 7 |
| 35 | RedChief, CI 12109 | 2 |
| 36 | RedChief x Pawnee | 7 |
| 37 | RedChief x Blackhull | 7 |
| 38 | RedChief x Nebred | 8 |
| 39 | RedChief x Marquillo-Oro | 7 |
| 40 | RedChief x Early Blackhull-Tenmarq | 7 |

Table 2 (concl.).

| Entry No., 1946-48 Tests | Kind | No. Selections Tested, 1951 and 1952 |
|---|---|---|
| 41 | Pawnee, CI 11669 | 2 |
| 42 | Pawnee x Blackhull | 7 |
| 43 | Pawnee x Nebred | 7 |
| 44 | Pawnee x Marquillo-Oro | 8 |
| 45 | Pawnee x Early Blackhull-Tenmarq | 7 |
| 46 | Blackhull, CI 6251 | 2 |
| 47 | Blackhull x Nebred | 7 |
| 48 | Blackhull x Marquillo-Oro | 7 |
| 49 | Blackhull x Early Blackhull-Tenmarq | 7 |
| 50 | Nebred, CI 10094 | 2 |
| 51 | Nebred x Marquillo-Oro | 7 |
| 52 | Nebred x Early Blackhull-Tenmarq | 7 |
| 53 | Marquillo-Oro, CI 11979 | 2 |
| 54 | Marquillo-Oro x Early Blackhull-Tenmarq | 8 |
| 55 | Early Blackhull-Tenmarq, Ks 2757 | 2 |

A F5 generation bulk yield nursery was grown in 1948, the seed coming from the 1947 F4 nursery. Notes recorded for both F4 trials and the one F5 trial are the same as those taken on the F3 generation tests.

In the fall of 1948 seed from each F4 generation bulk and parent was space-planted. Ten F5 plants of each cross and five plants of each parent were selected at random from this spaced material the following summer. Yield trials on the selected lines were expected to indicate whether or not the bulked population results had accurately predicted the probability of selecting good segregates from each bulk cross.

Single eight-foot plant rows were grown in 1950. Each row was harvested separately. The number of lines retained from each cross was reduced to seven or eight by discarding any line that failed to produce enough seed (approximately 60 grams) for a yield test the following year. In those cases where more than eight lines yielded enough seed for further testing the lowest yielding line, or lines, was discarded. Two randomly selected lines of each parent were retained. The number of lines tested from each kind is indicated in Table 2.

A 7x7x7 cubic lattice design was used to test the 343 selected lines in 1951. Two-row plots eight feet long were used. Recorded notes of agronomic interest include date of flowering, plant height, test weight, and yield.

The same design for these 343 lines was employed in 1952, the only difference being that four-row plots were used instead of two-rowed ones. The seed used was that harvested in 1951. Recorded notes include the same information as in the previous year.

## Plot and Record Techniques Employed

All eight-foot rows used in yield tests were planted to ten feet or more and trimmed to eight prior to harvest. When two-row plots were grown both rows were cut and threshed for yield; the two center rows of four-row plots were harvested. Nursery rows were 12 inches apart in all cases. Yield was determined in grams per eight square feet.

Plant height was measured as the average distance, in inches, from the ground to the tip of the spike, excluding awns, of a number of plants, or groups of plants, selected at random in the plot.

First head was recorded in 1946 as the date in May when one-half the heads had emerged from the boot. In all subsequent tests the first head note was replaced by a date of flowering note, which was recorded as the date in May when one-half the heads had started to bloom. The latter notation is considered somewhat more accurate in indicating the earliness of a strain since it is at the time of blooming, not emergence of the head from the boot, that fertilization takes place and development of the grain is initiated. Both values are measures of the relative earliness of a strain and, while not strictly comparable with 1947 and 1948 dates of flowering, the 1946 date of heading results were analyzed along with those of other years.

Test weight was determined in pounds per bushel, using standard pint or one-half pint test kettles, depending upon the amount of grain available for testing. The same test kettle and beam was used throughout the weighing of any one nursery.

## Statistical Methods Employed

The large numbers involved often made it difficult to find suitable testing space, either at the cereal breeding nursery or on the college agronomy farm, Manhattan, Kansas. It is hoped that proper and sufficient randomization and replication helped control the many sources of variation that arise in a study such as this.

Statistical analysis on the early generation material grown in randomized complete blocks was performed by the author using the methods of Snedecor (18) and Cochran and Cox (5). Analysis of variance for each characteristic in each year took the following form:

| Source | df |
|---|---|
| Replications | (r-1) |
| Kinds[1] | (k-1) |
| Error | (r-1)(k-1) |
| Total | rk-1 |

where r = number of replications and k = number of kinds. Snedecor's F was calculated as the ratio of variance of kinds, $s_k^2$, to variance of error, $s_e^2$.

---

[1] "Kinds" will be used to include both crosses (of which there are 45) and varieties (10) grown together in the tests reported.

Whenever the analysis of variance showed a significant difference between kinds the standard error of a kind mean, $s_{\bar{K}}$, was calculated as $\sqrt{s_{e}^{2}/r}$. A least significant difference, LSD, was also calculated; thus LSD $= \sqrt{2}\, t\, s_{\bar{K}}$, using t to represent Student's t at the five per cent level with degrees of freedom equal to the degrees of freedom of error variance used to calculate the standard error of a kind mean.

The LSD was applied to the arrayed kind means in an attempt to break the arrays into groups that might be of value for predictive purposes later on. Thus, whenever a difference as large as or larger than the LSD occurred between two adjacent means in the array it was possible to break the array at that point and say that the entries above the break were different, as a group, from the entries below the break.

Tukey (21) presented a test of means which often makes it possible to separate out significantly different treatments when the analysis of variance shows a significant difference between treatments but the LSD is not sensitive enough to indicate which ones differ. In any group of three or more means the most straggling mean from the general mean of the group may be tested by use of the formula

$$\frac{\left[(m_1-\bar{m})/s_m\right]-\left[(6/5)\log k\right]}{3\left[(1/4)+(1/n)\right]} = T$$

where $m_1$ = most straggling mean, $\bar{m}$ = grand mean of the group, k = number of means in the group, $s_m$ = standard deviation of a mean (also referred to as standard error of a mean), and n = degrees of freedom for $s_m$. T refers to the normal frequency distribution and its value at the five per cent level, 1.96, was used throughout. If a straggling mean was found to be non-significant (the value of $T \leq 1.96$) the array was assumed to be homogeneous. If the straggling mean was significant it was removed from the array and the most straggling mean of the remaining group tested; this process continued until a straggling mean was found to be non-significant. Those means that had been split off each end of the array were handled as significantly different groups, both different than the remaining group in the middle. In no case was it deemed necessary or useful to attempt to divide an array into more than three groups: high, low, and middle.

In those cases where it was believed possible to secure more information from a combined analysis of two or more year's results the general methods prescribed by Cochran and Cox (5) and Yates and Cochran (24) were followed. Before the combined analysis was completed a test to determine whether or not the error variances of the experiments being combined could be considered homogeneous was performed. When only two variances were involved [combining $F_4$ (1947) data with $F_4$ (1948)] Snedecor's test in which F is calculated as the quotient of the larger variance divided by the smaller was used. In all experiments reported herein the degrees of freedom for both numerator and denominator of this ratio was 216; tabular probabilities of F were doubled because only the upper portion of the F distribution was used.

When three experiments were to be combined [F3 (1946) with F3 (1947) and F3 (1948), for example] Bartlett's test of homogeneity was employed to determine the homogeneity of the error variances involved. Unequal degrees of freedom for error variances were always encountered and the test took the following form:

| Year | df | SS | SS/df | | (df)(log $s_e^2$) |
|---|---|---|---|---|---|
| 1946 | 486 | $SS_1$ | $s_{e1}^2$ | log $s_{e1}^2$ | 486(log $s_{e1}^2$) |
| 1947 | 486 | $SS_2$ | $s_{e2}^2$ | log $s_{e2}^2$ | 486(log $s_{e2}^2$) |
| 1948 | 216 | $SS_3$ | $s_{e3}^2$ | log $s_{e3}^2$ | 216(log $s_{e3}^2$) |
| | 1188 | SSS | | | S(df)(log $s_e^2$) |

$\bar{s}_e^2 =$ SSS/1188

Chi-square, 2 df = 2.3026 [1188(log $\bar{s}_e^2$)-S(df)(log $s_e^2$)]

Correlations between most years and between generations for each characteristic were calculated by the product-moment method. The correlation coefficient, r, was obtained from the formula $r = Sx_1x_2/\sqrt{Sx_1^2 \; Sx_2^2}$ where $Sx_1x_2 = SX_1X_2 - [(SX_1)(SX_2)]/n$ and $Sx^2 = SX^2 - [(SX)^2]/n$. Significance of a correlation coefficient was determined from Snedecor's "Correlation Coefficients at the 5% and 1% Levels of Significance."[1]

The test of significance of the difference between two correlation coefficients was accomplished by transforming r to

---

[1] George W. Snedecor, Statistical Methods, p. 149.

Fisher's z and then testing the difference between the two z's by the formula $t = (z_{r1}-z_{r2})/s_{zd}$, in which t represents Student's t at infinity and $s_{zd} = \sqrt{2/(n-3)}$.

Snedecor's weighted-z was used to test the hypothesis that certain groups of three correlation coefficients might be represented by a single correlation coefficient. The following form was used:

| Sample | df | r | z | Weighted z | Weighted $z^2$ |
|---|---|---|---|---|---|
| 1 | $n_1-3$ | $r_1$ | $z_1$ | $(n_1-3)z_1$ | $(n_1-3)z_1^2$ |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| y | $n_y-3$ | $r_y$ | $z_y$ | $(n_y-3)z_y$ | $(n_y-3)z_y^2$ |
| | $S(n-3)$ | | | $S(n-3)z$ | $S(n-3)z^2$ |

$\bar{z} = S(n-3)z/S(n-3)$

Chi-square, y-1 df $= S(n-3)z^2 - \bar{z}\left[S(n-3)z\right]$

In those cases where chi-square was small (probability greater than 0.05) it was concluded that the correlation coefficients of the samples being tested might all be estimates of the same population correlation coefficient; therefore, an average correlation coefficient of the samples was determined by transforming average z back to r.

Other special statistical methods and techniques applied to the early generation results will be discussed at the appropriate place in the discussion of experimental results.

The results of the cubic lattice trials were analyzed by the punched-card method at the Statistical Laboratory, Kansas State College.

## Weather Conditions Encountered

While no attempt has been made to relate weather conditions to the results obtained, it is important that consideration be given to the wide range of weather conditions during the five important years of this project: 1946, 1947, 1948, 1951, and 1952. Plant response to uncontrollable environmental influence is so complex that results of similar trials will vary from year to year. The only protection a plant breeder can provide for himself is to test his material over a sufficient number of years so that he feels he has randomly sampled the various conditions expected to be encountered in future years.

The 1946 test was seeded under favorable soil moisture conditions in the fall of 1945. Late spring rains revived the crop, which was beginning to suffer from drought; leaf rust infection was moderate. The weather was cool, but dry, during the heading and filling season.

High fall rainfall gave the 1947 tests a good start; a heavy snow cover provided protection from extremely low January temperatures. Cool temperatures and heavy spring rains up to ripening time were favorable for growth and development of the crop. Little damage from insects was noted; leaf rust infection was high.

The 1948 tests were seeded under dry soil conditions; fall emergence and growth was slow and uneven. Winter snows, late spring rains, and favorable growing temperatures during May and June brought the crop along rapidly. Insect and disease damage was light. The harvest of 1948 was hindered by excessive rainfall.

Soil moisture conditions were favorable for early top growth of the 1951 wheat tests, but dry weather late in the fall caused the wheat to enter winter in a rather weakened condition. Lack of winterkilling at Manhattan, coupled with high rainfall during the spring months, produced a large grain crop. Harvest was delayed by continued excessive rainfall, resulting in some shattering, lodging, and lowered test weights. Leaf rust infection was moderate.

The 1952 test was seeded in dry topsoil but high soil moisture reserves brought about moderate fall growth, which was resumed in the spring even though little winter or spring moisture fell. Practically no insect or disease damage was noted. High June temperatures caused the wheat to ripen rapidly; the early harvest was beset by high winds, which caused some shattering.

## EXPERIMENTAL RESULTS

### Early Generation Yield

<u>F3 Generation</u>. Bulk yield trials for the years 1946, 1947, and 1948 were analyzed separately. The F value and error variance for each experiment is included in Table 3. In 1946 and 1947 the F value of kinds exceeded the one per cent tabular value and in 1948 it was considered equal to the one per cent tabular value, a more exact statement being impossible because of the unavailability of an F distribution table showing 54 and

216 degrees of freedom for kind variance and error variance, respectively. The results would indicate that only once in 100 or more times would values of F as large as or larger than those obtained occur as a result of sampling from a population in which all kinds yielded the same; therefore, it can be stated that a highly significant difference between the mean yields of kinds of wheat grown in the trials was observed in 1946, 1947, and 1948. The standard errors of kind mean yields, along with the five per cent least significant differences, are presented in Table 3.

Kinds were arrayed according to their F3 generation bulk mean yield for each year. In 1946 the yields varied from 217.5 to 127.6 grams per eight-foot row, a range of 89.9 grams. The range in 1947 was 191.9-139.9 = 52.0 grams and in 1948 was 269.7-213.9 = 55.8 grams. In none of the three years was the calculated LSD effective in separating the array into two or more distinct groups since the difference between any two adjacent means was never equal to, nor greater than, the LSD. It seemed desirable that the arrays be subdivided, if possible, into groups representing at least high yielding and low yielding kinds; therefore, the methods of Tukey were applied to each year's data. By these methods the arrayed mean yields of 1946 were divided into three groups: high yielding, including entries 19, 42, and 48; low yielding, entries 3, 27, and 41; and a mid-group made up of all the remaining entries. The 1947 mean yields, when subjected

Table 3. Information supplied by the analysis of variance of yield for the generations and years specified.

| | F3 Generation Bulk | | | F4 Generation Bulk | | F5 Generation Bulk |
|---|---|---|---|---|---|---|
| | 1946 | 1947 | 1948 | 1947 | 1948 | 1948 |
| Degrees of freedom, replications | 9 | 9 | 4 | 4 | 4 | 4 |
| Degrees of freedom, kinds | 54 | 54 | 54 | 54 | 54 | 54 |
| Degrees of freedom, error | 486 | 486 | 216 | 216 | 216 | 215 |
| Mean square, kinds | 3179.720 | 1340.770 | 663.275 | 794.861 | 617.313 | 771.146 |
| Mean square, error | 579.914 | 494.762 | 411.608 | 550.691 | 430.894 | 381.437 |
| F, between kinds | 5.483** | 2.710** | 1.611** | 1.443* | 1.433* | 2.022** |
| Standard error of kind mean | 7.62 | 7.03 | 9.07 | 10.50 | 9.28 | 8.73 |
| Least significant difference at 5% level | 21.10 | 19.50 | 25.30 | 29.30 | 25.90 | 24.35 |

* Significant (Pr $\leq$ .05)

** Highly significant (Pr $\leq$ .01)

to the same methods, showed a low yielding group comprised of entries 35, 37, and 38 with all the rest of the entries remaining in the high yielding group. Only one mean yield was removed from the 1948 array by Tukey's methods; that was entry number 7, indicated to be high yielding.

When experiments are repeated over a number of years it is possible to get information on differences between years as well as interaction of treatments with years if the data are subjected to a combined analysis. One of the assumptions underlying the analysis of variance is that of homoscedasticity or equal error variances, Eisenhart (6). The exact consequences of failure to meet this and other assumptions are not known, Cochran (4). The effect of unequal error variances is considered serious because it disturbs the level of significance and decreases the sensitivity of F- or t- tests. As a rule the F-test tends to give too many significant results. Bartlett's test of homogeneity of variances was applied to the error variances of the three F3 generation bulk yield trials. Unadjusted chi-square, with two degrees of freedom, was found to equal 8.803. The probability of getting a value of chi-square this large or larger when sampling from a population in which all the variances are actually equal, is less than 0.02; therefore, the hypothesis that the three error variances are equal should be rejected.

Results of a combined analysis, when it is known that the pooled error mean square is made up of heterogeneous experimental errors, must be approached with caution. The following is the analysis of variance for the combined F3 generation bulk yield data, with an extra column for expected mean squares of interest in the analysis:

| Source | df | MS | EMS |
|---|---|---|---|
| Replications in years | 22 | 7163.000 | |
| Years | 2 | 490432.000 | $\sigma^2+8\sigma^2_{YK}+440\sigma^2_{Y}$ |
| Kinds | 54 | 2071.520 | $\sigma^2+8\sigma^2_{YK}+25\sigma^2_{K}$ |
| Y x K | 108 | 1556.120 | $\sigma^2+8\sigma^2_{YK}$ |
| Pooled error | 1188 | 514.478 | $\sigma^2$ |
| Total | 1374 | | |

The coefficient 25 in the foregoing EMS represents the sum of the number of observations made on each kind during the three years. The coefficients 8 and 440 were calculated by Snedecor's average number of observations, $k_0$, formula. This is $k_0 = [1/(n-1)][Sk-(Sk^2/Sk)]$. For Y x K, for instance, n-1 = 3-1 = 2 years; Sk = 10+10+5 = 25; and $Sk^2 = 10^2+10^2+5^2 = 225$. Substituting into the original formula one has $k_0 = 1/2[25-(225/25)] = 16/2 = 8$ observations as the average number of observations per year on kind 1, kind 2, ...., kind 55. Similarly, for years, one has $k_0 = 1/2[1375-(680625/1375)] = 440$ observations per year.

From the EMS it is evident that the first appropriate test of significance is $F = (\sigma^2+8\sigma^2_{yk})/\sigma^2 = 1556.12/514.478 = 3.025$. This value is greater than the approximate one per cent tabular value for F with 180 and 1188 degrees of freedom, 1.35, and could be called highly significant without hesitation if the pooled error variance was known to be made up of homogeneous variances. Since the variances were shown to be unequal by

Bartlett's test of homogeneity reference is made to Cochran and Cox (5) for further guidance.

These authors state that "if one experiment has a much higher error variance than any of the others, F will be distributed approximately as the tabular F with (t-1) and n' degrees of freedom, where t is the number of treatments and n' is the number of error degrees of freedom in the experiment with the high error variance."[1] An inspection of the test for homogeneity will reveal that the year 1948 has an error variance of 411.6 with 216 degrees of freedom while 1946 has an error variance of 579.9 and 1947 has an error variance of 494.8, each with 486 degrees of freedom. Clearly the year 1948 is contributing more to the pooled error variance, considering that it has less than half the degrees of freedom, than either of the other two years. Assuming that F in the combined analysis is distributed with 54 and 216 degrees of freedom, the tabular value of F at the one per cent level becomes approximately 1.61. It may be concluded, then, that the true one per cent level of significance for F lies somewhere between 1.35 and 1.61, both values considerably below the calculated F value of 3.025.

Another inspection of the EMS column will reveal that the appropriate test of significance for kinds becomes $F = (\sigma^2 + 8\sigma^2_{yk} + 25\sigma^2_k)/(\sigma^2 + 8\sigma^2_{yk})$, since the interaction between years and kinds

---

[1] William G. Cochran and Gertrude M. Cox, Experimental Designs, p. 400.

has been shown to be highly significant. This gives the value $F = 2071.52/1556.12 = 1.331$ with degrees of freedom 54 and 108. This is not significant when compared with the approximate five per cent tabular value, 1.47. It does not seem to be necessary to calculate a range of significance in this case since, according to Cochran and Cox (5), "when interactions are present, and especially if they are large, the F-test of the treatment mean square against the interaction mean square is much less disturbed."[1] With a highly significant interaction the effect of heterogeneity of error variances becomes slight when interaction mean square is used to test treatment mean square, as above.

The appropriate test of the significance of differences between years is $F = (\sigma^2+8\sigma^2_{yk}+440\sigma^2_y)/(\sigma^2+8\sigma^2_{yk}) = 490432/1556.12 = 315.163$, highly significant.

The final conclusions from the combined analysis of the F3 generation bulk yield trials would be that there is no difference between the mean yields of the 55 kinds of winter wheat grown over a period of three years, 1946-48, but that the kinds responded differently to the environment encountered during the three years. There was a highly significant difference between years.

When yearly kind means were ranked from 1 to 55 according to yield, 1 representing the highest yield, the inability of any one kind to maintain the same relative ranking from year to year became evident. Table 4 supports this statement by showing the yearly ranking of 15 kinds chosen because they represent the top

---

1 <u>Loc. cit.</u>

Table 4. The rank of 15 kinds of winter wheat according to yield during early generation bulk tests, 1946-48. Rank 1 is highest yielding; 55 lowest.

| | F3 Generation Bulk | | | F4 Generation Bulk | | | F5 Generation Bulk |
|---|---|---|---|---|---|---|---|
| Entry No. | 1946 | 1947 | 1948 | 1947 | 1948 | 2-Yr. Av. | 1948 |
| 19 | 1 | 8 | 49 | 41 | 6 | 20 | 5 |
| 48 | 2 | 45 | 11 | 32 | 13 | 22 | 24 |
| 42 | 3 | 7 | 8 | 51 | 36 | 49 | 18 |
| *55 | 4 | 5 | 47 | 2 | 8 | 3 | 20 |
| *46 | 5 | 48 | 38 | 47 | 46 | 54 | 37 |
| 7 | 26 | 41 | 1 | 38 | 38 | 37 | 34 |
| 21 | 27 | 37 | 15 | 18 | 45 | 32 | 16 |
| 10 | 28 | 4 | 3 | 45 | 18 | 34 | 4 |
| 31 | 29 | 52 | 53 | 55 | 28 | 55 | 38 |
| 32 | 30 | 43 | 24 | 1 | 48 | 15 | 21 |
| 23 | 51 | 11 | 40 | 4 | 33 | 14 | 50 |
| *50 | 52 | 39 | 10 | 43 | 15 | 27 | 27 |
| 3 | 53 | 16 | 13 | 24 | 42 | 29 | 52 |
| 27 | 54 | 10 | 9 | 3 | 1 | 1 | 14 |
| *41 | 55 | 24 | 48 | 39 | 11 | 24 | 30 |

* Parental variety.

5, the middle 5, and the lowest 5 in the 1946 yield trial. It will be noted that even the supposedly true breeding parental varieties included fail to maintain similar rankings, even in relation to each other. In this same respect, it should be made clear that the kinds grown over this period of three years represent the same genetic stock in all cases. The seed source for all three years was the same.

F4 Generation. Bulk yield trials for 1947 and 1948 were analyzed separately; pertinent information on the analysis is included in Table 3. In both years the F values exceed the five per cent level and approach the one per cent level but cannot be said to equal it. The F values, being significant, indicate that in each year there is a difference between kind yields.

The 1947 kind mean yield array has a range of 192.6-130.8 = 61.8 grams per eight-foot row; 1948 has a 261.3-209.5 = 51.8 gram range. The calculated LSD was not effective in dividing either array into yield groups; the methods of Tukey showed entry 31 to be significantly lower yielding than the other fifty-four entries in 1947 but were not able to separate out a single mean from the 1948 array.

A test of the homogeneity of F4 generation bulk yield error variances shows F, with 216 degrees of freedom for both numerator and denominator, to equal 1.278. This value approximates the tabular F value at what is now the ten per cent level; therefore, the variances may be said to be homogeneous.

The following is the analysis of variance for the combined F4 generation bulk yield data:

| Source | df | MS |
|---|---|---|
| Replications in years | 8 | 4181.380 |
| Years | 1 | 673645.000 |
| Kinds | 54 | 803.604 |
| Y x K | 54 | 608.570 |
| Pooled error | 432 | 490.793 |
| Total | 549 | |

$F_a = 608.57/490.793 = 1.240$
$F_b = 803.604/490.793 = 1.637^{**}$
$F_c = 673645/490.793 = 1372.564^{**}$

In the combined analysis a test of the year x kind interaction gave an insignificant F value. Non-significant interaction leaves pooled error mean square as the proper term for testing kinds and years in the combined analysis. Highly significant F's were found in both cases.

The combined analysis shows that there was a difference between F4 generation kind mean yields over the two year period, 1947-48; that the two years differed from each other; and that the kinds tended to respond the same each year. The standard error of a mean yield was 7.01 and the LSD was 19.50 grams per eight-foot row.

The arrayed mean yields of kinds averaged over a two-year period ranged from 223.6 to 185.5, a spread of 38.1 grams. Neither the LSD nor the methods of Tukey were effective in breaking the arrayed means into yield groups.

Inspection shows that the F4 generation bulk kinds were somewhat more consistent in yield from year to year than the F3 generation kinds during the same two years. Table 4 shows the rank of 15 kinds of early generation bulks according to yield. It should be noted that the ranks given under "2-Year Average, F4 Generation" are not averages of ranks but are ranks of average yield.

<u>F5 Generation</u>. Analysis of the bulk yield trial for 1948 is summarized in Table 3. A highly significant difference between kinds was observed. The yield of entry 17 in the fifth replication was missing and had to be calculated by the missing-plot technique of Cochran and Cox (5). As a result one degree of freedom was lost from the analysis.

The arrayed mean yields ran from 241.4 to 184.3, a range of 57.1 grams. The LSD was not effective in breaking the array into yield groups; Tukey's methods were not applied to the means.

Comparisons of the yield rank of 15 of the 55 F5 generation bulk kinds with the rank of the same 15 kinds during F3 and F4 generations may be made from Table 4.

Early Generation Average Yield. Combined analysis of F3, F4, and F5 generation bulk and parent yields in 1946, 1947, and 1948, respectively, is presented because such an analysis represents the sequence of generations and years most likely to be used in a practical plant breeding program. It would not always be practical nor possible to grow the same generation during two or three successive years, as was done for this study.

Bartlett's test of homogeneity shows that chi-square, with two degrees of freedom, equals 12.639. The probability of getting a chi-square this large or larger from a population in which all the variances are actually equal is less than 0.01. The variances for F3 (1946), F4 (1947), and F5 (1948) cannot be considered homogeneous. The error variance of the F4 generation test, with 216 degrees of freedom, is considerably higher than the other two.

In the combined analysis the effect of generations cannot be separated from the effect of years because only one generation per year is being considered. The following analysis of variance applies:

| Source | df | MS |
|---|---|---|
| Replications in years | 17 | 5950.620 |
| Years | 2 | 141561.000 |
| Kinds | 54 | 1965.030 |
| Y x K | 108 | 1390.350 |
| Pooled error | 918 | 525.922 |
| Total | 1099 | |

$F_a = 1390.35/525.922 = 2.644^{**}$
$F_b = 1965.03/1390.35 = 1.413$
$F_c = 141561/1390.35 = 102.817^{**}$

The F value for year x kind interaction is above the one per cent point of significance for either 108 and 918 or 54 and 216 degrees of freedom. Interaction mean square, being highly significant, becomes the proper error term for testing kinds and years. The F for kinds in this combined analysis was non-significant; the differences between years were highly significant. The heterogeneous error variances involved in this combined analysis have little influence on the last two tests because interaction, not pooled error, mean square has been shown to be the proper denominator for the F-tests.

Conclusions from the combined analysis would indicate that there is a difference between years (or generations), no difference between the average yields of these 55 early generation kinds over the three-year period, and a tendency for kinds to respond differently from year to year.

## Early Generation Plant Height

F3 Generation. Plant height of F3 generation bulk kinds for the years 1946, 1947, and 1948 was analyzed separately; pertinent parts of each analysis are included in Table 5. In the 1947 test the height of entry 31 in replication IV was missing; one degree of freedom was lost from total and error in the analysis when the value was calculated by the missing-plot technique of Cochran and Cox (5). In all cases the calculated F for kinds was highly significant. Interpretation suggests that unless a 1 in 100 chance has come off there is a difference in plant height between kinds in each of the three years.

The array of F3 generation kinds according to mean plant height ran from 41.7 to 36.5, a spread of 5.2 inches, in 1946. The range was 45.7-40.8 = 4.9 inches in 1947 and 41.4-35.2 = 6.2 inches in 1948. In none of the three trials was the LSD effective in separating the array into significantly different height groups. The methods of Tukey split the 1946 array into three groups: tall, entries 29, 46, and 53; short, entries 10, 11, 19, 27, 54, and 55; and a medium height group made up of the remaining forty-six entries. Tukey's methods served only to separate out entries 15, 20, and 41 as short kinds in 1947; the remaining entries were shown to belong to one group. The 1948 array was divided by Tukey's methods so that entries 28, 29, and 53 fell into the tall group; entries 6, 11, 12, 15, 17, 19, 32, 41, 43, 44, 45, 51, 52, 54, and 55 formed the short group; and thirty-seven entries remained in the medium height group. It

Table 5. Information supplied by the analysis of variance of plant height for the generations and years specified.

| | F3 Generation Bulk | | | F4 Generation Bulk | | F5 Generation Bulk |
|---|---|---|---|---|---|---|
| | 1946 | 1947 | 1948 | 1947 | 1948 | 1948 |
| Degrees of freedom, replications | 9 | 9 | 4 | 4 | 4 | 4 |
| Degrees of freedom, kinds | 54 | 54 | 54 | 54 | 54 | 54 |
| Degrees of freedom, error | 486 | 485 | 216 | 216 | 216 | 216 |
| Mean square, kinds | 10.380 | 12.280 | 9.501 | 8.279 | 11.690 | 11.900 |
| Mean square, error | 1.457 | 3.932 | 1.094 | 3.458 | 1.819 | 2.071 |
| F, between kinds | 7.122** | 3.124** | 8.681** | 2.394** | 6.426** | 5.745** |
| Standard error of kind mean | 0.382 | 0.627 | 0.468 | 0.832 | 0.603 | 0.644 |
| Least significant difference at 5% level | 1.06 | 2.63 | 1.30 | 2.30 | 1.70 | 1.80 |

** Highly significant (Pr ≤ .01)

should be noted that the 1946 and 1948 tall groups both include entries 29 and 53. Four of the six short entries in 1946 are in the short group of 1948. Entries 15 and 41 were identified as short kinds in 1947 and 1948.

Bartlett's test of homogeneity shows the error variances for the three individual tests to be non-homogeneous; the 1947 test, with 485 degrees of freedom, has the highest error variance.

The combined analysis of variance follows:

| Source | df | MS |
|---|---|---|
| Replications in years | 22 | 85.720 |
| Years | 2 | 3868.000 |
| Kinds | 54 | 24.760 |
| Y x K | 108 | 3.701 |
| Pooled error | 1187 | 2.402 |
| Total | 1373 | |

$F_a = 3.701/2.402 = 1.541^*$
$F_b = 24.76/3.701 = 6.690^{**}$

The test of year x kind interaction by pooled error results in an F value which is highly significant with 108 and 1187 degrees of freedom and is significant with 54 and 485 degrees of freedom. The significant year x kind interaction becomes the proper term for testing the significance of differences between kinds and between years. The F value for kinds in this combined analysis is highly significant. The difference between years can be shown to be highly significant by inspection of the mean square column above.

The combined analysis shows that there is a difference between the mean heights of the 55 F3 generation kinds; there is a year to year difference in the heights observed; and that the kinds do not respond the same from year to year as far as height is concerned. The standard error of a kind mean height when taken over the three-year period was 0.385 inch. The five per cent LSD was 1.05 inches.

Arrayed mean heights for the 55 kinds grown over a three-year period had a range of 42.9-38.9 = 4.0 inches. There was no difference between adjacent means as large as the LSD. Tukey's methods break the array into three groups: tall, made up of entries 4, 5, 7, 24, 28, 29, 35, 37, 46, and 53; short, entries 15, 41, and 52; and a medium height group made up of the remaining forty-two entries. All entries that appeared in the tall group of 1946 or 1948 are included in the tall group of average heights.

Inspection of Table 6 will reveal that F3 generation kinds are fairly consistent in rank according to height from year to year. The 15 entries included in Table 6 were chosen because they were the 5 tallest, the 5 shortest, and 5 mid-height kinds in the 1946 F3 generation bulk test. In general, the tall kinds of any one year will be found among the tall kinds of any other year; practically the same relationship is evident for the short kinds. Those kinds with rankings mid-way between tall and short tend to fluctuate in rank somewhat from year to year, but generally do not overlap with tall or short kinds. The bulk hybrids are apparently not more inconsistent in height from year to year than the parental varieties included.

Table 6. The rank of 15 kinds of winter wheat according to plant height during early generation bulk tests 1946-48. Rank 1 is tallest; 55 shortest.

| Entry No. | F3 Generation Bulk 1946 | 1947 | 1948 | 3-Yr. Av. | F4 Generation Bulk 1947 | 1948 | 2-Yr. Av. | F5 Generation Bulk 1948 | Av. of the 3 Generations Grown in Successive Years |
|---|---|---|---|---|---|---|---|---|---|
| *46 | 1 | 1 | 9 | 1 | 1 | 7 | 1 | 4 | 1 |
| 29 | 2 | 4 | 3 | 2 | 3 | 6 | 4 | 6 | 2 |
| *53 | 3 | 22 | 1 | 5 | 23 | 1 | 6 | 1 | 3 |
| 7 | 4 | 5 | 5 | 4 | 8 | 15 | 12 | 7 | 7 |
| *28 | 5 | 2 | 2 | 3 | 6 | 2 | 2 | 5 | 5 |
| 14 | 26 | 6 | 36 | 15 | 44 | 37 | 44 | 33 | 38 |
| 26 | 27 | 31 | 33 | 31 | 14 | 35 | 23 | 16 | 16 |
| 8 | 28 | 19 | 17 | 24 | 30 | 20 | 29 | 12 | 26 |
| 13 | 29 | 17 | 22 | 25 | 16 | 27 | 19 | 14 | 19 |
| 42 | 30 | 39 | 31 | 33 | 42 | 40 | 43 | 38 | 45 |
| 54 | 51 | 44 | 45 | 49 | 52 | 48 | 52 | 42 | 51 |
| 10 | 52 | 26 | 35 | 41 | 32 | 24 | 31 | 20 | 37 |
| 27 | 53 | 37 | 29 | 43 | 54 | 28 | 48 | 22 | 46 |
| 19 | 54 | 27 | 51 | 51 | 41 | 54 | 50 | 50 | 50 |
| *55 | 55 | 47 | 55 | 46 | 51 | 55 | 55 | 55 | 55 |

* Parental variety

<u>F4 Generation</u>. Plant height of F4 generation bulk kinds in 1947 and 1948 was analyzed in a manner identical to the F3 generation analyses. Table 5 shows important parts of each analysis. Highly significant F values for kinds in both years indicate that there was a difference in mean plant height of kinds each year. LSD at the five per cent level was 2.3 and 1.7 inches in 1947 and 1948, respectively.

Arrayed plant heights for the F4 generation bulks grown in 1947 run from 46.4 to 40.6, a spread of 5.8 inches; the range of plant heights for the 1948 trial is 42.0-35.0 = 7.0 inches. The LSD did not indicate a significant difference between any pair of adjacent mean heights in either year. When Tukey's methods were applied to the 1947 array entry 41 was shown to be short; the other fifty-four entries could not be separated. The 1948 array, when subjected to the same methods, divided into a tall group consisting of entries 3, 4, 20, 24, 28, 29, 30, 31, 32, 46, 47, 48, and 53 and a medium height group consisting of all the remaining entries except 55, which fell into the short group alone.

An F value of 1.901 with 216 degrees of freedom in both numerator and denominator shows that unless a less than 1 in 50 chance has come off the variances of the two F4 generation tests are non-homogeneous. The 1947 error variance is considerably higher than the 1948 error variance; both have 216 degrees of freedom.

The combined analysis takes the following form:

| Source | df | MS |
|---|---|---|
| Replications in years | 8 | 20.820 |
| Years | 1 | 3333.000 |
| Kinds | 54 | 15.830 |
| Y x K | 54 | 4.135 |
| Pooled error | 432 | 2.638 |
| Total | 549 | |

$F_a = 4.135/2.638 = 1.567^{**}$
$F_b = 15.83/4.135 = 3.829^{**}$
$F_c = 3333/4.135 = 806.119^{**}$

The F value for the test of significance of year x kind interaction is highly significant with the 54 and 432 degrees of freedom that would apply if error variances were homogeneous and also remains highly significant if 54 and 216 degrees of freedom, representing the higher of the heterogeneous error variances, are used. The proper error term for testing kinds and years, then, is the year x kind interaction mean square. The F value for kinds is highly significant. The test for years also indicates a highly significant difference. The standard error of a kind mean height in the two-year average is 0.643 inch. The five per cent LSD is 1.94 inches.

It is safe to conclude from the combined analysis that there was a difference between 1947 and 1948 as far as plant height of the 55 kinds is concerned; that there was a difference between mean plant heights of the 55 kinds averaged over the two-year period; and that the kinds did not respond the same from year to year.

The array of F4 generation plant heights averaged over the two-year period runs from 43.6 to 38.3, a range of 5.3 inches. The LSD was not effective in breaking the array into height groups but Tukey's method shows that entries 4, 21, 24, 28, 29, 32, 46, and 53 may be considered a tall group; 41, 43, and 55 a short group; and the remaining forty-four entries a medium height group. All of the entries that fall into the tall group in the two-year average analysis, except 21, were shown to belong to the tall group in 1948.

The rank, according to height, of 15 F4 generation bulks and parents is included in Table 6. In general, the agreement in rank of any particular kind from year to year is good. Entry 53, the Marquillo-Oro parent, was in twenty-second and twenty-third place in the F3 (1947) and F4 (1947) tests, respectively; however, it climbed to first place in both 1948 tests.

F5 Generation. Plant height for the F5 generation bulks and parents grown in 1948 was subjected to an analysis of variance, important parts of which are included in Table 5. Again the differences in height between kinds were found to be significant.

The arrayed mean heights range from 42.4 to 34.8, a spread of 7.6 inches. The LSD did not indicate a single mean in the array to be significantly different from means adjacent to it. The Tukeyized data fell into three plant height groups: tall, made up of entries 4, 24, 46, and 53; short, consisting of 15, 41, 44, 45, and 55; and a medium height group consisting of forty-six entries.

Table 6 shows the ranking of 15 of the entries according to plant height.

Early Generation Average Height. Combined analysis of F3, F4, and F5 generation plant heights in 1946, 1947, and 1948, respectively, is presented for the reasons previously mentioned. Bartlett's test of homogeneity shows that the three error variances are non-homogeneous (probability less than 0.01); the F4 (1947) error variance with 216 degrees of freedom, was considerably higher than the other two.

The form for the analysis of variance of the combined experiments is as follows:

| Source | df | MS |
|---|---|---|
| Replications in years | 17 | 78.106 |
| Years | 2 | 2167.000 |
| Kinds | 54 | 24.200 |
| Y x K | 108 | 3.176 |
| Pooled error | 918 | 2.072 |
| Total | 1099 | |

$F_a = 3.176/2.072 = 1.533^*$
$F_b = 24.2/3.176 = 7.621^{**}$

F for year x kind interaction is highly significant at 108 and 918 degrees of freedom and significant at 54 and 216 degrees of freedom. The latter degrees of freedom must be considered in this analysis because the error variances that were pooled to make up pooled error variance were not homogeneous. The significant interaction mean square becomes the proper error term for testing kinds and years. The F value for kinds is highly significant. It can be determined by inspection that there is a highly significant difference between years.

Conclusions to be drawn from the combined analysis are that there is a difference in plant height from year to year (or from generation to generation); that there is a difference between three-year average plant heights of kinds; and that individual kinds may differ in height from year to year (or from generation to generation). The standard error of a kind mean height in this analysis is 0.398 inch. The LSD at the five per cent level is 1.18 inches.

The arrayed mean heights for combined F3 (1946), F4 (1947), and F5 (1948) generation tests have a range of 43.1-37.6 = 5.5 inches. The LSD did not divide the array into height groups; Tukey's methods were not applied to the array.

Table 6 includes a column showing the rank of the average height of 15 kinds grown as F3 bulks in 1946, F4 bulks in 1947, and as F5 bulks in 1948. The kinds are apparently fairly consistent in height from generation to generation and from year to year. It seems that the tall and the short kinds retain their relative rank somewhat better than the medium height kinds.

The tendency, also mentioned in connection with F3 generation results, for greater variability in rank to occur among the medium height group than among the tall or short groups may have a genetic explanation. The inheritance of plant height in wheat is complex but it seems safe to say that it is influenced by many factor pairs. It is suggested that some crosses may have a preponderance of factors for tallness or shortness while others, the intermediate types, will have nearer equal numbers of genes for tallness and shortness. These latter types might vary in their response to different year to year environmental changes according to the actual number or proportion of each kind of gene they carry. Conversely, types that are largely tall or short genotypically might show a relatively consistent height response, even under changing year to year environment.

The four short hybrids in Table 6 involve Early Blackhull-Tenmarq as a common parent. This short variety evidently is effective in transmitting its genes for shortness to hybrids in which it is involved.

## Early Generation Date of Flowering

F3 Generation. Summarized analyses of the date of flowering of F3 generation bulks and parents for the years 1946, 1947, and 1948 are presented in Table 7. A highly significant difference in flowering date between kinds was observed each year. Only once in 100 or more times would F values as large as or larger than those obtained be encountered when sampling from a population in which the date of flowering was actually all the same. The standard errors of kind means and five per cent LSD values, in days, for each year are included in Table 7.

Yearly arrays of kinds according to flowering date, where May 1 equals 1, have ranges of 11.7-(-2.8) = 14.5 days in 1946, 23.7-16.4 = 7.3 days in 1947, and 23.6-14.6 = 9.0 days in 1948. The LSD, when applied to the 1946 array, showed entry 53 to be late; entries 10, 19, 27, 34, 40, 45, 49, 52, 54, and 55 to be early; and the remaining entries to belong to a mid-maturity group. The LSD was not effective in dividing the 1947 and 1948 F3 generation arrays into maturity groups. Tukey's methods show that the 1947 array can be divided into a late group, made up of entries 20, 24, 25, 50, 51, and 53; an early group, comprised of entries 2, 6, 10, 11, 13, 14, 15, 16, 17, 18, 19, 27, 30, 34, 36, 40, 41, 42, 43, 44, 45, 49, 52, 54, and 55; and a mid-maturity group made up of the remaining twenty-four entries. The same methods divide the 1948 array into a late group, entries 1, 9, 20, 22, 24, 25, 47, 50, and 53; an early group, entries 10, 13, 14, 16, 19, 27, 34, 36, 40, 45, 49, 52, 54, and 55; and a mid-maturity group made up of thirty-two entries. Five of the six late entries in 1947 were again classified as late in 1948. All of the early entries in 1946 and 1948 were included in the long list of entries in the early group of 1947.

The error variances of the three F3 generation experiments were shown to be non-homogeneous by Bartlett's test. The error variance for the 1948 experiment, with 216 degrees of freedom,

Table 7. Information supplied by the analysis of variance of date of flowering for the generations and years specified.

| | F3 Generation Bulk | | | F4 Generation Bulk | | F5 Generation Bulk |
|---|---|---|---|---|---|---|
| | 1946 | 1947 | 1948 | 1947 | 1948 | 1948 |
| Degrees of freedom, replications | 9 | 9 | 4 | 4 | 4 | 4 |
| Degrees of freedom, kinds | 54 | 54 | 54 | 54 | 54 | 54 |
| Degrees of freedom, error | 486 | 486 | 216 | 216 | 216 | 216 |
| Mean square, kinds | 92.810 | 30.600 | 21.900 | 13.900 | 30.510 | 24.400 |
| Mean square, error | 1.358 | 0.890 | 0.752 | 0.897 | 1.437 | 0.692 |
| F, between kinds | 68.353** | 34.406** | 29.082** | 15.506** | 21.226** | 35.228** |
| Standard error of kind mean | 0.368 | 0.298 | 0.388 | 0.424 | 0.536 | 0.372 |
| Least significant difference at 5% level | 1.00 | 0.80 | 1.10 | 1.20 | 1.50 | 1.00 |

** Highly significant ($Pr \leq .01$)

was high compared to the other two variances. The combined analysis of variance was as follows:

| Source | df | MS |
|---|---|---|
| Replications in years | 22 | 9.183 |
| Years | 2 | 34930.000 |
| Kinds | 54 | 133.240 |
| Y x K | 108 | 6.033 |
| Pooled error | 1188 | 1.056 |
| Total | 1374 | |

$F_a = 6.033/1.056 = 5.712^{**}$
$F_b = 133.24/6.033 = 22.083^{**}$

The F value for year x kind interaction is well above the one per cent range of significance, which would include 1.35 and 1.41, allowing for heterogeneous error variances. The F value for kinds, when tested by year x kind interaction mean square, is also highly significant. It is evident, from inspection of the mean square column above, that the F value for years is highly significant.

Interpretation of the combined analysis shows that there was a difference between the years 1946, 1947, and 1948 as far as flowering date of the 55 F3 generation bulks and parents is concerned; that there was a difference between the flowering dates of kinds averaged over the three-year period; and that the kinds did not all react the same from year to year.

The arrayed mean dates of flowering for the three-year period have a range of 18.9-8.4 = 10.5 days. The LSD between adjacent means in this array is 1.4 days. There is not this large a difference between any pair of adjacent means in the array. Tukey's methods were not applied to the data.

The ranking of kinds according to the date upon which one-half the heads had initiated bloom shows that the F3 generation kinds were fairly consistent in relative rank from year to year. Table 8 shows the rank of 15 early generation kinds during all generations and years tested. The 15 kinds were chosen because they were the 5 latest, the 5 earliest, and 5 of the mid-maturing kinds in the F3 generation test of 1946. Most noteworthy is the way the 5 early kinds continue to be the 5 earliest kinds in all three years; the 5 late kinds do not deviate much from being the latest in all three years. Apparently environmental changes did not operate to change the relative date of flowering of the F3 generation bulks from year to year. The column headed "3-Year Average" in Table 8 gives the rank of the same 15 kinds when their dates of flowering in all three F3 generation experiments are averaged.

<u>F4 Generation</u>. Summarized analyses of the 1947 and 1948 F4 generation dates of flowering are included in Table 7. F values for the difference between kind dates of flowering were significant in both experiments.

Arrayed F4 generation kind dates of flowering ran from 23.2 to 16.8, a spread of 6.4 days, in 1947. The 1948 range is 25.0-11.4 = 13.6 days. Neither the LSD nor the methods of Tukey were effective in breaking the array of 1947 mean flowering dates into maturity groups. In 1948 the LSD showed entry 53 to be late; entries 10, 19, 27, 34, 40, 45, 49, 52, 54, and 55 to be early; and left the remaining forty-four entries in a mid-maturity group.

When the two F4 generation error variances were tested they were found to be non-homogeneous. The 1948 error variance was greatest; the degrees of freedom for either was 216.

Table 8. The rank of 15 kinds of winter wheat according to date of flowering during early generation bulk tests, 1946-48. Rank 1 is latest; 55 earliest.

| | F3 Generation Bulk | | | | F4 Generation Bulk | | | F5 Generation Bulk | Av. of the 3 Generations Grown in Successive Years |
|---|---|---|---|---|---|---|---|---|---|
| Entry No. | 1946 | 1947 | 1948 | 3-Yr. Av. | 1947 | 1948 | 2-Yr. Av. | 1948 | |
| *53 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *20 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 2 |
| *50 | 3 | 5 | 3 | 3 | 5 | 5 | 4 | 4 | 3 |
| 25 | 4 | 3 | 4 | 4 | 6 | 2 | 3 | 2 | 4 |
| *46 | 5 | 8 | 11 | 6 | 10 | 8 | 10 | 11 | 8 |
| 3 | 26 | 20 | 25 | 25 | 27 | 22 | 25 | 12 | 24 |
| 7 | 27 | 26 | 28 | 27 | 29 | 26 | 29 | 26 | 30 |
| 8 | 28 | 27 | 23 | 28 | 19 | 30 | 26 | 29 | 28 |
| *35 | 29 | 29 | 40 | 29 | 43 | 40 | 43 | 42 | 36 |
| 6 | 30 | 34 | 27 | 31 | 33 | 25 | 31 | 17 | 29 |
| 10 | 51 | 51 | 51 | 51 | 47 | 50 | 50 | 50 | 50 |
| 45 | 52 | 54 | 55 | 54 | 52 | 52 | 52 | 54 | 53 |
| 19 | 53 | 53 | 52 | 53 | 48 | 53 | 53 | 51 | 52 |
| 34 | 54 | 52 | 53 | 52 | 54 | 54 | 54 | 52 | 54 |
| *55 | 55 | 55 | 54 | 55 | 55 | 55 | 55 | 55 | 55 |

* Parental variety

The combined analysis was completed as follows:

| Source | df | MS |
|---|---|---|
| Replications in years | 8 | 4.363 |
| Years | 1 | 4.540 |
| Kinds | 54 | 40.900 |
| Y x K | 54 | 3.549 |
| Pooled error | 432 | 1.167 |
| Total | 549 | |

$F_a = 3.549/1.167 = 3.040^{**}$
$F_b = 40.9/3.549 = 11.516^{**}$
$F_c = 4.54/3.549 = 1.279$

The above analysis shows that there is a highly significant year x kind interaction, even with 54 and 216 degrees of freedom to allow for heterogeneity of error variances; therefore, interaction mean square becomes the proper error term for testing kinds and years. The second F-test shows that there is a highly significant difference between kind dates of flowering averaged over the two years. There is no significant difference between years. The standard error of an average flowering date is 0.596; the LSD at the five per cent level is 1.7 days.

The array of kind flowering dates averaged over a two year period runs from 24.1 to 14.1, a spread of 10 days. The LSD did not split the array into maturity groups; the methods of Tukey were not applied.

The ranks, according to date of flowering, of 15 kinds of winter wheat grown in the F4 generation experiments of 1947 and 1948 are shown in Table 8. The third column under "F4 Generation Bulks" of this table gives the rank of average date of flowering over a two-year period for the same 15 kinds. The rank for F4 generation kinds remains fairly constant from year to year and the ranks of each generation grown in any one year do not deviate greatly. The three-year average of F3 generation dates of flowering corresponds quite well with the two-year average of F4 generation dates of flowering.

F5 Generation. The analysis of the 1948 results of the date of flowering study on F5 generation bulks and parents is shown in Table 7. A highly significant difference between kind dates of flowering was observed.

The array of F5 generation dates of flowering has a range of 26.2-15.2 = 11.0 days. The LSD at five per cent showed entry 53 to be significantly later than the remainder of the array and entries 10, 19, 27, 34, 40, 45, 49, 52, 54, and 55 to be earlier. This left forty-four entries in a mid-maturity group. It should be noted that the maturity groups set up for the F5 generation test of 1948 include exactly the same entries as the corresponding groups for the F4 generation test and are included in the corresponding groups of the F3 generation test that same year.

Table 8 shows the rank, according to date of flowering, of 15 representative kinds grown in the F5 generation bulk test. The agreement in rank from generation to generation, except for entries 3 and 6, is quite close when looking at the 1948 results; the agreement from year to year is close with the same two exceptions. The two entries noted bloomed earlier, in relation to other kinds included, as F5 generation bulk hybrids in 1948 than they had in any other generation or year.

Early Generation Average Flowering Date. As previously mentioned, it is believed that the most practical way to obtain and use whatever information of value early generation bulk testing may provide is to grow the successive generations in successive years. For this reason a combined analysis of flowering date of F3 generation bulks and parents in 1946, F4 generation bulks and parents in 1947, and F5 generation bulks and parents in 1948 was undertaken.

Bartlett's test of homogeneity applied to the error variances of the experiments in question gives a chi-square, with two degrees of freedom, of 12.604. The probability of getting a chi-square that large or larger when the variances are actually all homogeneous is less than 0.01. The error variances that make up the pooled error should not be considered homogeneous. The $F_4$ (1947) error variance, with 216 degrees of freedom, is higher than the other two.

The effect of years and generations cannot be separated and whatever effect each may have is combined under "years" in the following analysis:

| Source | df | MS |
|---|---|---|
| Replications in years | 17 | 11.957 |
| Years | 2 | 2946.000 |
| Kinds | 54 | 118.980 |
| Y x K | 108 | 6.057 |
| Pooled error | 918 | 1.084 |
| Total | 1099 | |

$F_a = 6.057/1.084 = 5.587^{**}$
$F_b = 118.98/6.057 = 19.645^{**}$

In the first test of significance the F value for year x kind interaction is greater than the one per cent limits of significance set up by 108 and 918 or 54 and 216 degrees of freedom. The second test, utilizing interaction mean square as the error term, gives a highly significant F for the differences between kinds. Inspection of the mean square column is all that is necessary to show the highly significant difference between years. Besides the difference between years and between kinds averaged over three successive generation-years, it may be concluded that the relationship between kind flowering dates does not remain the same from year to year.

The standard error of a kind mean was 0.591; the five per cent LSD was 1.66 days. Arrayed mean dates of flowering of the 55 kinds averaged over three successive generation-years ranged from 20.4

to 9.7, a spread of 10.7 days. The LSD showed entry 53 to be late and left the other fifty-four entries in one group. Tukey's methods were not employed.

The last column of Table 8 shows the rank of the average heading date for 15 kinds taken over three successive generation-years. The agreement of ranks in this average with the ranks of individual generation averages is close. The upper and lower ends of the flowering date rank distributions are in close agreement while the ranks of mid-maturity kinds deviate somewhat but still retain the same general position.

The apparent greater variation in rank of mid-maturity kinds over the early or late kinds may be explained in a manner similar to that for variation in intermediate height groups. It is suggested that the early and late types have a predominance of factors for earliness or lateness, whichever the case may be, and that these types will not be greatly influenced in relative rank of date of flowering by year to year environmental changes. The mid-maturity types would respond to year to year environmental changes somewhat more, if the same line of reasoning is followed, because their genotype is made up of nearer equal humbers of factors for earliness and lateness; the response would be conditioned by the number or proportion of each kind of factor present.

The four early bulk hybrids in Table 8 involve Early Blackhull-Tenmarq as a common parent. This variety, the earliest of the parents studied, must effectively transmit its genes for earliness to hybrids involving it.

Table 9. Information supplied by the analysis of variance of test weight for the generations specified, 1948.

| | Generation | | |
|---|---|---|---|
| | F3 Bulk | F4 Bulk | F5 Bulk |
| Degrees of freedom, replications | 1 | 1 | 1 |
| Degrees of freedom, kinds | 54 | 54 | 54 |
| Degrees of freedom, error | 54 | 54 | 54 |
| Mean square, kinds | 2.170 | 1.960 | 2.390 |
| Mean square, error | 0.249 | 0.110 | 0.116 |
| F, between kinds | 8.751** | 17.811** | 20.614** |
| Standard error of kind mean | 0.353 | 0.235 | 0.241 |
| Least significant difference at 5% level | 1.00 | 0.67 | 0.68 |

** Highly significant (Pr ≤ .01)

## Early Generation Test Weight

<u>F3 Generation</u>. Table 9 summarizes the only single-year analysis applicable to test weight of the F3 generation bulks and parents. Only one observation on test weight of grain produced by the F3 generation kinds was made in 1946 and in 1947; two observations were made in 1948. The F value for 1948 kinds was highly significant, indicating that there was a true difference between kind test weights.

The array of F3 generation test weights in 1948 ran from 59.4 to 54.0, a range of 5.4 pounds. The LSD indicated that entry 35 could be considered heavier in test weight than the other fifty-four entries.

The range of test weights in the 1946 experiment was 63.6-56.8 = 6.8 pounds; in 1947 the range was 61.0-55.8 = 5.2 pounds per bushel. No tests of significance are applicable to the results of either of these years; therefore, statistical tests of combined data will not be as powerful as tests of other characteristics previously discussed. The values obtained from a single observation are used as the best estimate of test weight available for that generation and year.

The best that can be done in the way of a combined F3 generation analysis seems to be to consider each of the total four sets of observations over the three-year period as a "replication" and analyze accordingly. No test of error variance homogeneity is possible. The proposed analysis of variance takes the following form:

| Source | df | MS |
|---|---|---|
| "Replications" | 3 | 278.000 |
| Kinds | 54 | 3.840 |
| Error | 162 | 0.599 |
| Total | 219 | |

$F = 3.84/0.599 = 6.411^{**}$

The highly significant F for kinds indicates that there is a difference in test weight between the 55 kinds. The standard error of a kind mean test weight is 0.387; the LSD at five per cent is 1.08 pounds.

An array of test weights of F3 generation kinds averaged over the three-year period runs from 60.3 to 55.2, a spread of 5.1 pounds. Differences as large as the LSD do not appear between any two adjacent means in the array. Tukey's methods were not employed.

The rank of 15 representative kinds according to test weight is recorded in Table 10. The kinds were chosen because they were the 5 heaviest, the 5 lightest, and 5 medium weight kinds in the

1946 F3 generation experiment. Certain kinds appear to be much more consistent from year to year in relative test weight rank than other kinds. The bulk hybrids are apparently not more variable in rank than the parental varieties.

F4 Generation. Only one observation of test weight was recorded for the F4 generation bulks and parents grown in 1947; two observations were recorded in 1948, analysis of which is summarized in Table 9. A highly significant difference between test weights of kinds is indicated.

The array of test weights recorded for the 1948 F4 generation experiment has a range of 60.0-54.6 = 5.4 pounds. The LSD showed entry 35 to be heavy, entry 20 to be light, and left the remaining fifty-three entries in one group. Arrayed test weights for 1947, when only one observation was made, ranged from 62.0 to 58.0, a spread of 4.0 pounds.

Tests of homogeneity of error variances do not apply. The combined analysis uses each of the three sets of observations as a "replication" and takes the following form:

| Source | df | MS |
|---|---|---|
| "Replications" | 2 | 146.000 |
| Kinds | 54 | 2.270 |
| Error | 108 | 0.247 |
| Total | 164 | |

$F = 2.27/0.247 = 9.201^{**}$

There is a difference between test weights of kinds, as indicated by the highly significant F value. The standard error of a kind mean is 0.287; the five per cent LSD is 0.81 pound.

Arrayed F4 generation kind mean test weights averaged over the two-year period have a range of 60.7-56.3 = 4.4 pounds. The

LSD, when applied to this array, shows only that entry 35 is a heavy kind.

Ranks, according to test weight, of F4 generation kinds are presented in Table 10. Agreement from year to year is not particularly close; nor is the rank of average F3 generation test weight in close agreement with average F4 generation test weight. In general, it seems that the lighter kinds tend to hold their relative rank from year to year better than other kinds.

F5 Generation. Two observations of test weight were made on the F5 generation bulk experiment. Table 9 records pertinent parts of the analysis of variance. Unless a 1 in 100 chance has come off, there is a difference between test weights of the 55 kinds tested.

The arrayed test weights of F5 generation kinds had a range of 58.2-52.4 = 5.8 pounds. The five per cent LSD showed entry 35 to be significantly higher than the rest of the entries. Entry 35, the RedChief parent, has previously been shown to be heavier than any other entry in either the F3 or F4 generation tests.

The rank, according to test weight, of 15 kinds is shown in Table 10. Again, the rather close year to year and generation to generation agreement in relative rank of the lighter kinds is apparent.

Early Generation Average Test Weight. In order to obtain a clearer picture of the results that might be obtained from early generation test weight trials in a practical breeding program,

Table 10. The rank of 15 kinds of winter wheat according to test weight during early generation bulk tests, 1946-48. Rank 1 is heaviest; 55 lightest.

| Entry No. | F3 Generation Bulk 1946 | 1947 | 1948 | 3-Yr. Av. | F4 Generation Bulk 1947 | 1948 | 2-Yr. Av. | F5 Generation Bulk 1948 | Av. of the 3 Generations Grown in Successive Years |
|---|---|---|---|---|---|---|---|---|---|
| *55 | 1 | 16 | 19 | 5 | 21 | 2 | 9 | 21 | 6 |
| 40 | 2 | 1 | 3 | 2 | 8 | 3 | 5 | 3 | 2 |
| 49 | 3 | 7 | 5 | 3 | 4 | 7 | 3 | 8 | 3 |
| 19 | 4 | 20 | 35 | 16 | 25 | 19 | 21 | 16 | 7 |
| 45 | 5 | 39 | 50 | 34 | 33 | 28 | 29 | 48 | 16 |
| *46 | 26 | 19 | 13 | 18 | 51 | 12 | 22 | 19 | 33 |
| 31 | 27 | 31 | 14 | 22 | 28 | 14 | 18 | 12 | 22 |
| 5 | 28 | 26 | 23 | 24 | 5 | 18 | 11 | 22 | 17 |
| 13 | 29 | 27 | 21 | 25 | 24 | 24 | 25 | 15 | 26 |
| 16 | 30 | 28 | 10 | 21 | 10 | 21 | 17 | 20 | 21 |
| 26 | 51 | 52 | 46 | 50 | 50 | 53 | 53 | 51 | 52 |
| * 1 | 52 | 54 | 53 | 54 | 45 | 52 | 51 | 52 | 53 |
| 44 | 53 | 53 | 49 | 53 | 54 | 51 | 52 | 53 | 54 |
| 18 | 54 | 51 | 30 | 46 | 49 | 49 | 50 | 45 | 51 |
| *53 | 55 | 55 | 55 | 55 | 55 | 54 | 55 | 55 | 55 |

* Parental variety

an attempt to analyze successive generations in successive years was made. Here, again, the single observations in 1946 and 1947 reduce the weight of conclusions that can be drawn from the analysis. The analysis, considering each set of observations to be a "replication," was as follows:

| Source | df | MS |
|---|---|---|
| "Replications" | 3 | 547.000 |
| Kinds | 54 | 3.920 |
| Error | 162 | 0.528 |
| Total | 219 | |

$F = 3.92/0.528 = 7.422^{**}$

The highly significant F value for kinds indicates that there is a significant difference between test weights of kinds averaged over the three years. The standard error of a kind mean test weight is 0.364; the five per cent LSD is 1.02 pounds.

The arrayed average test weights of kinds grown for three successive generations run from 60.7 to 55.7, a range of 5.0 pounds. The LSD does not separate the array into test weight groups.

The last column of Table 10 shows the rank of 15 kinds of wheat according to their average test weight over three successive generations grown in as many successive years. Test weight is a quantitative character presumably conditioned by many genes. A possible genetic explanation of the apparent tendency for the lightest kinds in the F3 generation test of 1946 to remain at the lower end of the test weight rank distribution in all subsequent generations and years lies in the interaction of these multiple factors. Types that are heavy or light would be expected to have a predominance of factors for heaviness or lightness, respectively; those that are intermediate in test weight would have nearer equal numbers of the two kinds of factors. A type that is genotypically heavy or light would tend to remain relatively more stable under different environments than intermediate types, which would respond according to the number or proportion of each kind of gene present.

Application of this explanation to the results shown in Table 10 would lead to the conclusion that the genotypes of the five lightest kinds in the 1946 F3 generation test were made up of a predominance of factors for low test weight; therefore, the response of these kinds remained stable in all generations and years. The other ten kinds in Table 10, however, must have had nearer equal numbers of factors for heavy and light; relative response of these kinds varied from year to year because of environmental influence on the interaction of these genes.

It seems that the F3 generation test in 1946 effectively pointed out the kinds that would be low in test weight in succeeding generations and years but that the same test was of little value for showing the high test weight kinds. The three hybrids in the low test weight group of Table 10 involve Marquillo-Oro, a low test weight variety, as a common parent. Evidently Marquillo-Oro effectively transmits its low test weight to crosses involving it.

## Correlations in Early Generations

To facilitate a more complete study of the relationship between generations and between years for any one character, product-moment correlations for many combinations of early generations and years were calculated. Table 11 summarizes these correlations.

Yield. Further support for the lack of year to year consistency in yield of the F3 generation bulks and parents is provided by a study of the correlations calculated for all possible combinations of the three years involved. In none of the three years did the correlation coefficient, r, approach significance.

Table 11. Product-moment correlation coefficients of early generation bulk data, 1946-48.

| | Yield | Plant Height | Date of Flowering | Test Weight |
|---|---|---|---|---|
| | r = | r = | r = | r = |
| F3 (1946) vs F3 (1947) | 0.0017 | 0.8492** | 0.9584** | 0.8439** |
| F3 (1947) vs F3 (1948) | 0.0143 | 0.8405** | 0.9502** | 0.6894** |
| F3 (1946) vs F3 (1948) | 0.1463 | 0.9065** | 0.9824** | 0.4828** |
| Average, F3 (1946) to F3 (1948) | 0.0540 | 0.8690** | -- | -- |
| F4 (1947) vs F4 (1948) | 0.1392 | 0.5945** | 0.9058** | 0.6634** |
| F3 (1946) vs F4 (1947) | -0.1134 | 0.7138** | 0.9371** | 0.5370** |
| F4 (1947) vs F5 (1948) | -0.0897 | 0.6782** | 0.9027** | 0.6992** |
| F3 (1946) vs F5 (1948) | 0.1186 | 0.7542** | 0.9293** | 0.6511** |
| Average, F3 (1946) to F5 (1948) | -- | 0.7280** | 0.9245** | 0.6630** |
| F3 (1947) vs F4 (1947) | 0.3965** | 0.9567** | 0.9313** | 0.6979** |
| F3 (1948) vs F4 (1948) | -0.0241 | 0.8792** | 0.9665** | 0.8826** |
| F4 (1948) vs F5 (1948) | 0.2276 | 0.8279** | 0.9536** | 0.8504** |
| F3 (1948) vs F5 (1948) | 0.1276 | 0.9408** | 0.9513** | 0.9301** |

** Highly significant ($Pr \leq .01$)

All r's are based on 55 paired comparisons.
r at 5%, 53 df, is 0.266; at 1%, 53 df, it is 0.345.

The three yield correlation coefficients were tested to see if they might all be estimates of a single coefficient of correlation. The weighted-z test of Snedecor, previously described, gave a chi-square of 0.6639 with two degrees of freedom. The probability of getting a chi-square this large or larger is approximately 0.71. There was no reason to reject the null-hypothesis; therefore, the three correlation coefficients were assumed not to be unlike. The average yield correlation coefficient of the 55 kinds over a period of three years was calculated to be r = 0.054, which is non-significant.

The low yield correlations between F3 generations grown in successive years would indicate that the F3 kinds vary in their yield response from year to year; this is in agreement with the highly significant year x kind interaction reported in the analysis of variance of the combined F3 generation tests. Apparently the yield of F3 generation kinds in any one year was not indicative of the yield of the same F3 generation kinds in any other year tested.

Although the correlation of yield of F4 generation bulks and parents in 1947 and 1948 is higher than the F3 generation correlation for the same years, the F4 generation correlation coefficient of 0.1392 does not approach significance. The relative yield of F4 generation kinds in 1948 was not predicted by the yield of the same F4 generation kinds in the previous year. The correlation coefficient indicates that the F4 generation kinds varied in relative yield from year to year but the analysis of variance of combined F4 generation results showed a non-significant year x kind interaction.

An explanation of this apparent discrepancy probably lies in the sensitivity of the analysis of variance as applied to the combined F4 generation yield results, although the possibility that random sampling variation is contributing a few values that are out of line cannot be entirely ruled out. In the combined analysis the pooled error used for testing year x kind interaction is actually the second order interaction "year x kind x replications in years." Kind x replications in years interaction cannot be isolated in this analysis; it is considered likely that the interaction of kinds with both years and replications is sufficient to increase the size of the denominator of the appropriate F-test to a point where year x kind interaction is indicated to be non-significant. An inspection of Fig. 1, showing a scatter diagram of the 55 F4 generation kind yields in 1947 plotted against their F4 generation yields in 1948, supports the contention that there is no relationship between F4 generation yields in the two years tested.

Further study of other correlations brings out the doubtful value of trying to determine the yield of bulked crosses during early generations. Yield correlations between all possible combinations of the F3 generation grown in 1946, the F4 generation grown in 1947, and the F5 generation grown in 1948 are reported in Table 11. These combinations were chosen because they represent the most likely sequence of three year's data that could be chosen from this study. None of the correlation coefficients approach significance.

It became a problem of interest to determine whether or not the three correlation coefficients -0.1134, -0.0897, and 0.1186 might all be considered equal and representative of a single population correlation coefficient. If the r's were found to be dissimilar, it was reasoned, there might be some basis for saying that yield was more highly correlated between certain pairs of

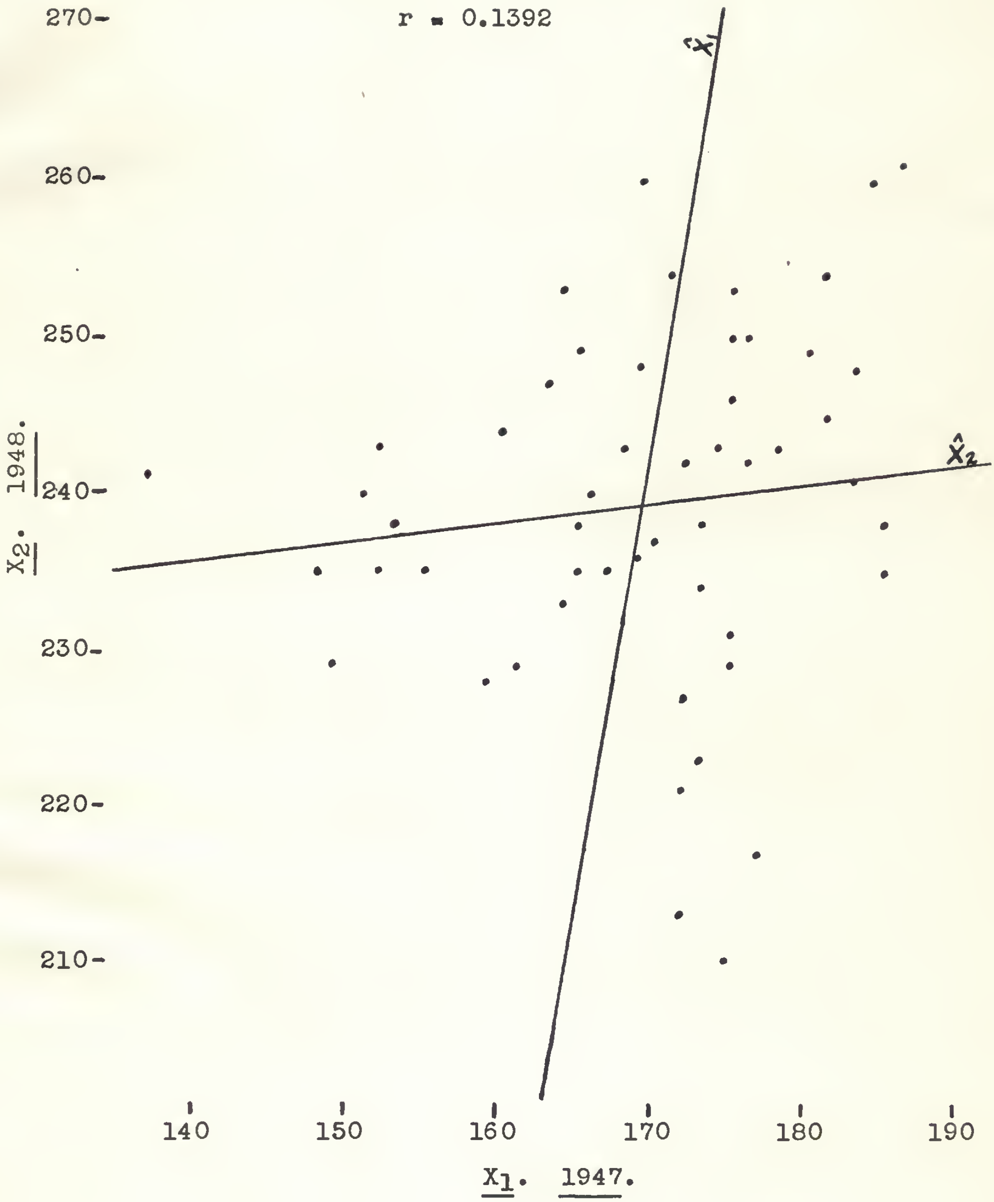


Figure 1. Yield, in grams per 8-foot row, of 55 kinds of winter wheat grown in F4 generation bulk tests.

generations than between other pairs. Snedecor's weighted-z shows that the three r's can all be considered equal; therefore, the ability of tests during one generation to predict the yield of future generations has not been established. The low correlation between any two of the early generations was equal to the correlation of any other pair of early generations grown in three successive years.

Other tests of yield correlations were made to determine whether the correlation coefficients for successive generations in successive years differed from the coefficient of correlation for the same set of years when genetically identical stocks were grown. This might give an indication, it was postulated, of whether variation in yield correlations was caused by generation differences or was brought about principally by year to year environmental influences. Tests of the difference between the following sets of correlations were made by the appropriate method:

1. F3 (1946) vs F4 (1947) and F3 (1946) vs F3 (1947). Results: t at infinity equal to 0.5867; probability is over 0.50.

2. F3 (1946) vs F5 (1948) and F3 (1946) vs F3 (1948). Results: t at infinity equal to 0.1429; probability is over 0.50.

3. F4 (1947) vs F5 (1948), F3 (1947) vs F3 (1948), and F4 (1947) vs F4 (1948). Results: chi-square, with 2 df, equal to 0.7073; probability is approximately 0.73.

In all cases the hypothesis that the r's were the same was accepted and it was concluded that the r's involved in each test could be considered equal. Apparently the yield correlation between generations, already shown to be negligible, is as high as the correlation between yields of genetically identical plant material grown in successive years.

A further study of yield correlations involved the elimination of year to year environmental influences by correlating successive generations grown in one year. The highly significant F3 (1947) vs F4 (1947) correlation is probably due to sampling variation; it will be noted that the correlation between the same two generations in 1948 is negative and non-significant. Yield determinations on bulks and parents in any one early generation do not seem to be indicative of the yield of any other generation grown the same year.

A review of the yield data collected on early generation bulk hybrids and parents shows that it is impossible to isolate any high yielding or low yielding individual kind or group of kinds. It would seem foolhardy to base selection of crosses to be retained in the breeding program on the results of any one or any logical combination of early generation yield tests described herein.

Plant Height. Reference is again made to Table 11, showing product-moment height correlations for all possible yearly combinations of the F3 generation tests. All the correlations are highly significant, indicating that the kinds, as a whole, tend to remain fairly constant in plant height from year to year. The weighted-z test shows that all three r's may be considered estimates of the same population correlation coefficient; the average r for the F3 generation height over the three-year period 1946 to 1948 is 0.869, which is highly significant.

The correlation of 1947 with 1948 plant height of the 55 F4 generation kinds was 0.5945, a value which is somewhat below those recorded for F3 generation height correlations but is still highly significant.

In both cases, above, there is apparently some disagreement between conclusions to be drawn from the analysis of variance and those indicated by a study of correlations. The combined analysis of F3 generation tests and of F4 generation tests showed significant or highly significant year x kind interactions; such interaction is not apparent in the correlations. It should be understood that the analysis of variance will generally show a significant year x kind interaction whenever the difference between two treatments varies significantly from year to year. This would include both the situation in which there is no change in rank and the one in which the ranks of treatments change from year to year. Simple correlations indicate the degree of linear relationship between two variables; the magnitude of differences between treatments loses its importance and significance is obtained whenever most of the differences between treatments have been in the same direction. The apparent discrepancy in results of the two approaches, then, may be due to a situation in which the difference between kinds, or some of the kinds, is significantly different from year to year but the rank of kinds according to plant height remains similar during the same years.

Further explanation of this apparent discrepancy between conclusions drawn from the analysis of variance and correlation studies will be attempted under the heading "Date of Flowering."

Table 11 also reports height correlations calculated for all possible combinations of F3, F4, and F5 generations grown in successive years. All these correlations are highly significant and show that the height of plants in any one generation and year is fairly indicative of the height of the same kinds in other generations and years. The r values for successive generation-years are somewhat lower than the r values for those same years when the correlations of the F3 generation only were concerned. This indicates a situation in which there is some generation to generation variation added to the expected year to year variation.

Snedecor's weighted-z shows that the three r's for successive generation-years may be considered equal. Chi-square, with two degrees of freedom, is 0.772 and the probability of getting a chi-square as large as or larger than this value is 0.68 when sampling from a normal population. The average correlation coefficient for successive generation-years is 0.728, a highly significant r. The height of plants in any one generation apparently was linearly related to the height in any other generation.

To determine whether or not generations had an effect upon plant height correlations the following tests involving comparisons between height correlations when different generations were grown in different years and when genetically identical material was grown in the corresponding years were made:

1. F3 (1946) vs F4 (1947) and F3 (1946) vs F3 (1947).
   Results: t at infinity equals 1.8367; probability is 0.07.

2. F3 (1946) vs F5 (1948) and F3 (1946) vs F3 (1948).
   Results: t at infinity equals 1.2398; probability is 0.22.

3. F4 (1947) vs F5 (1948), F3 (1947) vs F3 (1948), and F4 (1947) vs F4 (1948).
   Results: chi-square, with 2 df, equals 8.2615; probability is less than 0.01.

The r's in each of the first two tests may be considered equal. Although the probability in the first test is approaching significance it is still reasonable to assume that the correlation between F3 and F4 generation plant height in 1946 and 1947, respectively, is as high as the correlation between heights observed in 1946 and 1947 when the F3 generation only is concerned. A similar statement holds true for the second test.

The third test, with a chi-square probability of less than 0.01, indicates that the three r's tested cannot be considered equal. A break-down of the weighted-z test into individual comparisons shows that the F4 (1947) vs F5 (1948) correlation coefficient is not different from either the F3 (1947) vs F3 (1948) or the F4 (1947) vs F4 (1948) correlation coefficients, but that the F3 (1947) vs F3 (1948) correlation coefficient does differ from the one for F4 (1947) vs F4 (1948). Conclusions are rather hard to draw, but it seems that the relative height of most kinds in one generation gives a reliable indication of the relative height of the same kinds in succeeding generations.

Successive generations grown in the same year were correlated in order to eliminate year to year environmental influence on early generation plant height correlations. All correlations were highly significant and were somewhat higher numerically than corresponding correlations of the same generations grown in different years. This lends support to the earlier statement regarding year x kind interaction as it affects plant height. The effect of years definitely must be present; however, the effect is generally small and these results indicate that it may be safe to use the plant height of a bulk or parent in any one generation as a criterion of the performance expected from that same bulk or parent in a later generation, whether it be grown the same year or some other year.

It has been shown that the bulk hybrids and parents were fairly consistent in relative height from generation to generation and from year to year. It might be safe to base selection for plant height in such a group on the results of any one test reported; average results of a combination of successive generation tests in successive years would be more accurate because the best estimate of plant height of those kinds affected by differential year to year response would become the average over a random sample of years.

Date of Flowering. Table 11 shows product-moment correlations of flowering date for all possible combinations of years in which the F3 generation experiments were grown. The correlation coefficients are highly significant and are numerically high, indicating a near perfect agreement from year to year. Snedecor's weighted-z test showed that the three correlation coefficients should not all be considered estimates of the same population correlation coefficient, however. When individual comparisons between r's were made it was found that the 1946 vs 1947 correlation coefficient did not differ from either the 1947 vs 1948 or the 1946 vs 1948 correlation coefficients, but that the 1947 vs 1948 r could not be considered equal to the 1946 vs 1948 r. This finding might seem unexpected at first glance but is understandable when one considers that it takes small numerical differences between such high correlations to cause the weighted-z test to show a significant difference; the same numerical difference between two low correlations would not approach significance.

The correlation between dates of flowering in 1947 and 1948 for the 55 F4 generation bulks and parents is 0.9058, a highly significant correlation coefficient.

Again there is an apparent discrepancy between conclusions to be drawn from the combined analyses of variance and a correlation study of F3 or F4 generation results. The combined analyses of variance show highly significant year x kind interactions for each generation; the high correlations would indicate that the response of kinds remains the same from year to year. There is actually no conflict between the two conclusions; significant year x kind interactions indicate the failure of the difference between two kind dates of flowering to remain practically constant from year to year; such interactions are possible without any change in rank, whatsoever. The correlations, meanwhile, show that there was a close linear relationship between dates of flowering of kinds from year to year; the magnitude of difference between kinds has little effect.

In an attempt to explain any possible conflict between conclusions drawn from the two approaches a scatter diagram, Fig. 2, of F4 generation dates of flowering in 1947 plotted against F4 generation dates of flowering in 1948 was drawn up. It shows that the points are well distributed along the linear trend lines and that the correlation is not due to two or three clusters of points. It will be observed that a few points show some deviation from the trend lines but that the bulk of the points show little deviation. The value for entry 55 at 16.8, 11.4 deviates most widely from the linear trend lines.

55 individual, non-orthogonal, year x kind interaction mean squares were calculated in order to locate the source of the year x kind interaction in the combined analysis of F4 generation dates of flowering. Each individual interaction mean square, representing each of the 55 kinds involved, was tested against pooled error. By this analysis 17 entries out of the 55 were shown to have a significant or highly significant year x kind interaction. Entry 55 contributed far more to combined interaction mean square than any other entry. It is realized that the completion of so many non-orthogonal comparisons opens such an analysis to immediate criticism since the probability level for some of the tests of significance is upset. Presentation is made merely to show

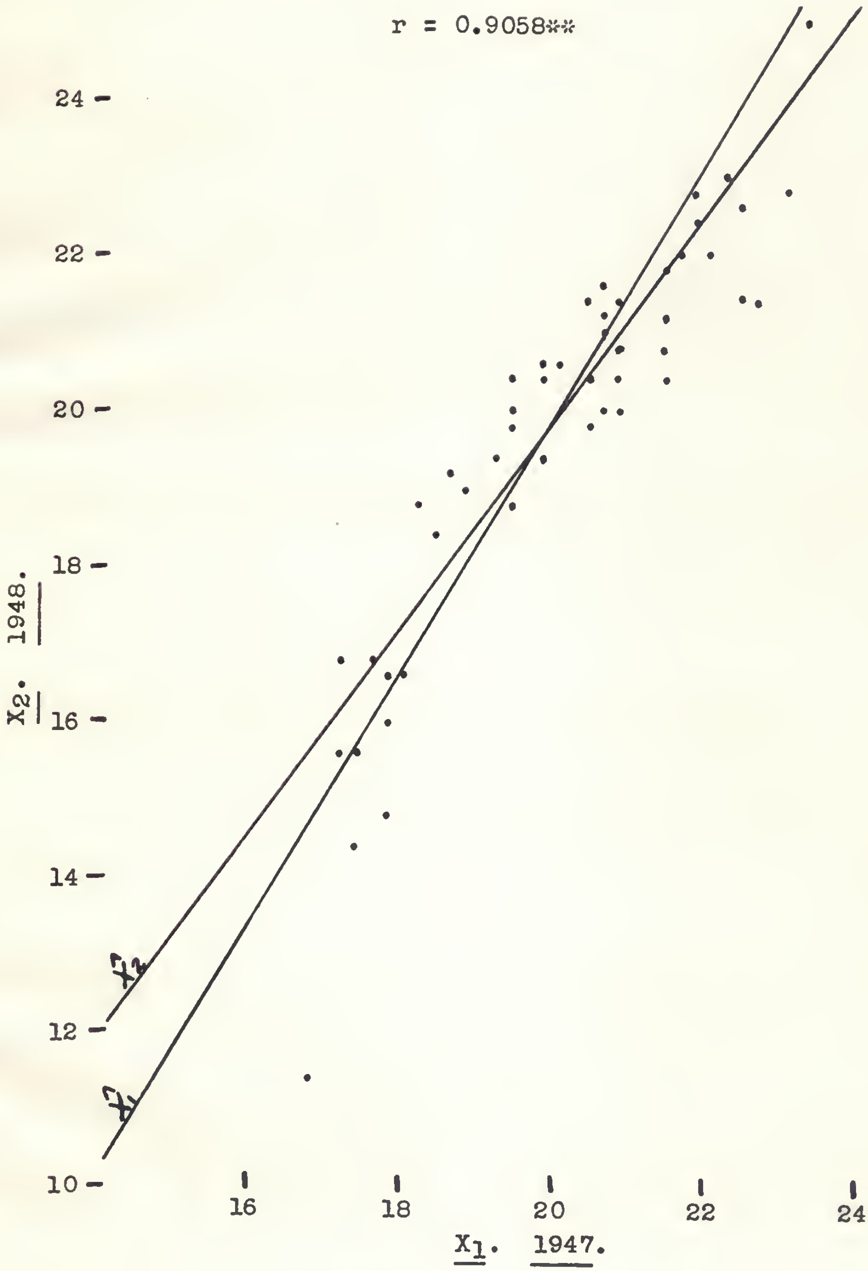


Figure 2. Date of flowering, in days after May 1, of 55 kinds of winter wheat grown in F4 generation bulk tests.

that the year x kind interaction in the combined analysis is probably the result of the failure of only a few of the 55 F4 generation entries to respond the same in 1947 and 1948. If this assumption is correct the high correlation of the remaining entries may be obscuring the effect of these few variable entries, thereby causing the entire group to be apparently correlated with an r value of 0.9058.

The two explanations of the apparent discrepancy between conclusions to be drawn from an analysis of variance and from a correlation study of the same data do not oppose each other. It seems reasonable to believe that both operate at the same time. The explanation concerning interaction due to a few entries being counterbalanced by high correlation among the remaining entries is probably of lesser importance, however, and would apply when significant year x kind interactions are the result of changes in rank.

The correlation coefficients for all possible combinations of F3, F4, and F5 generations grown in successive years are recorded in Table 11. All three of these correlations are highly significant and their high numerical value gives added weight to the conclusion that the flowering date of kinds in one generation is indicative of the date those same kinds will bloom in other generations.

The weighted-z test of Snedecor was used to test the hypothesis that the three correlation coefficients are estimates of the same population correlation coefficient. A chi-square of 1.382, with two degrees of freedom, has a probability of approximately 0.50. The correlation coefficients may be considered equal and an average correlation coefficient for date of flowering between successive generation-years is found to be approximately 0.9245, a highly significant r.

The following tests of correlation coefficients for generations and years specified compared with correlation coefficients for genetically identical material grown in the same specified years were made:

1. F3 (1946) vs F4(1947) and F3 (1946) vs F3 (1947).
   Results: t at infinity equal to 1.0765; probability is approximately 0.28.

2. F3 (1946) vs F5 (1948) and F3 (1946) vs F3 (1948).
   Results: t at infinity equal to 3.6122; probability is less than 0.01.

3. F4 (1947) vs F5 (1948), F3 (1947) vs F3 (1948), and F4 (1947) vs F4 (1948).
   Results: chi-square, with 2 df, equals 3.9581; probability is approximately 0.16.

The r's in the first test may be considered equal, indicating that the correlation of flowering date between F3 and F4 generations grown in 1946 and 1947, respectively, is as good as that between 1946 and 1947 when the F3 generation only is considered.

The second test shows that the r's involved cannot be considered equal. The correlation of flowering date between F3 and F5 generations grown in 1946 and 1948, respectively, is not the same as that between 1946 and 1948 when the F3 generation only is considered.

The third test shows that all date of flowering correlation coefficients obtained between 1947 and 1948 are equal, whether different generations were grown in the two years or whether the same generation was involved in both years.

To complete the date of flowering correlation study successive generations grown in the same year were correlated. This has the effect of removing the influence year to year variation would have on the inter-generation correlations. As a whole the correlation between generations is the same whether they were grown the same year or in different years.

It has been shown that it is possible to arrange the early generation bulk hybrids and parents into groups according to date of flowering in any one test and that this grouping will reliably indicate the performance that can be expected from entries in the group in any succeeding generation or year. An average of the date of flowering results over two or more generations grown in successive years would be even more reliable as an indication of early generation performance because such an average would cover a random sample of years and thus take into account the year to year variation shown to affect some kinds.

Test Weight. Product-moment correlations of test weight between all possible combinations of the F3 generation tests grown over the three-year period are recorded in Table 11. All correlations are highly significant but appear to vary from year to year. The weighted-z test shows that the three r's cannot be considered equal. Individual comparisons of correlation coefficients reveal that the 1947 vs 1948 and 1946 vs 1948 r's may be considered equal, but that the 1946 vs 1947 r differs from that of the other two correlations.

The correlation between F4 generation test weight in 1947 and 1948 was 0.6634, a highly significant r.

The correlation of test weight between all possible combinations of F3 (1946), F4 (1947), and F5 (1948) generations are recorded in Table 11. A weighted-z test shows that all three r's may be considered equal. The correlation between any two of the successive generation-years is apparently as great as that between any other pair of generation-years that might be chosen. The average correlation coefficient for F3, F4, and F5 generations, grown in that order in successive years, is 0.663, highly significant.

Tests to determine whether test weight correlation coefficients determined for certain generations and years differed from correlation coefficients for the same years when genetically identical plant material was grown were conducted as follows:

1. F3 (1946) vs F4 (1947) and F3 (1946) vs F3 (1947).
   Results: t at infinity equal to 3.2398; probability is less than 0.01.

2. F3 (1946) vs F5 (1948) and F3 (1946) vs F3 (1948).
   Results: t at infinity equal to 1.2653; probability is approximately 0.21.

3. F4 (1947) vs F5 (1948), F3 (1947) vs F3 (1948), and F4 (1947) vs F4 (1948).
   Results: chi-square, with 2 df, equal to 0.1162; probability is approximately 0.94.

In the first test the correlation between F3 and F4 generation test weights in 1946 and 1947, respectively, was not as high as that between 1946 and 1947 when the F3 generation only was considered. It will be remembered that the r value for F3 (1946) vs F3 (1947) was shown to be not equal to (actually greater than) that of any other two-year combination involving only the F3 generation.

The second test indicates that the correlation between test weights of the F3 and F5 generations grown in 1946 and 1948, respectively, does not differ from that of 1946 and 1948 when only the F3 generation is considered.

The third test indicates that none of the three 1947-48 correlation coefficients differ, whether the effect that generations might possibly have on test weight is accounted for or not.

Test weights of successive generations grown in one year were correlated in an attempt to determine whether or not the effect of years had any influence on the correlations. In all cases the correlation between successive generations grown in one year was greater than for the same generations grown in successive years. It would seem that the effect of years does operate to reduce inter-generation correlations of test weight.

A satisfactory grouping of early generation bulk hybrids and parents according to test weight can probably be made by averaging the test weights of two or more generations grown in successive years. Any determinations made on one generation grown in any one year may prove disappointing if used to characterize the expected performance of a kind in some later generation or year. Although there is a certain amount of consistency in test weight from generation to generation and year to year the relationship is not as strong as that reported for plant height and date of flowering. The test weight of any other generation grown that same year can be reliably predicted from the results of a single generation test, but it is difficult to see what the practical value of such a prediction would be.

## Tests of Selections

A cubic lattice design was employed to test the 343 random selections from 45 crosses and ten parental varieties in 1951 and 1952. Relative efficiency of a cubic lattice indicates the amount of information secured over that which would have been secured by analyzing the same data as randomized complete blocks.

Yield. Highly significant differences in yield were found to exist among the 343 entries grown in 1951 and 1952. The range of 1951 adjusted yields was 246.2-112.4 = 133.8 grams per eight-foot row; the five per cent LSD for comparing differences between any two adjusted mean yields was 28.90 grams. In 1952 the range was 233.2-110.4 = 122.8 grams per eight-foot row; the LSD was 32.80 grams. Relative efficiency of the cubic lattice design compared to a randomized complete block was 119.5 per cent in 1951 and 135.5 per cent in 1952.

Since it had previously been shown that it was impossible to make a reliable determination of the yield of early generation bulk hybrids grown in this experiment it seemed unnecessary to make such a detailed study of the yield of selections as will be reported for other characteristics of the same selections. The yield of selections was undoubtedly subject to similar year to year variation as that encountered when testing bulk hybrids; therefore, it would not have been safe to designate any selection or group of selections as high or low yielding on the basis of one or two year's testing.

In 1951 the LSD indicated that there was a significant difference in yield between the two selections from Marquillo-Oro, entry 53. In no other case during the two-year period was there a significant difference between the two selections from each parent. This indicates that, in general, there was more variation between the yield of parental varieties than between the two selections from any variety. It appeared that the best available estimate of the yield performance of selections from each cross or variety was the mean yield of all selections from each cross or variety averaged over the two years in which they were grown. When averaged in this manner the yield of all selections from each kind had a range of 198.3-160.4 = 37.9 grams per eight-foot row. It was found that the selections from Marquillo-Oro, entry 53, had the highest average yield; those from Pawnee x Early Blackhull-Tenmarq, entry 45, the lowest. These entries had not been identified as high or low yielding kinds during early generation bulk tests.

Plant Height. There was a significant difference in plant height among the 343 selections tested in 1951 and 1952. Adjusted plant heights ran from 41.4 to 32.2, a spread of 9.2 inches, in 1951; the five per cent LSD for comparing differences between any two mean heights was 0.79 inch. In 1952 the range was 49.0-35.6 = 13.4 inches; the LSD was 2.14 inches. One of the selections from entry 15, Comanche x Pawnee, was the shortest line in both years. Entry 15 was consistently among the short kinds in early generation tests.

Relative efficiencies of the cubic lattice design compared to randomized complete blocks were 402 per cent in 1951 and 173 per cent in 1952. The efficiency for 1951 seems out of line when compared to other efficiencies, but checks on the computation reveal no error. No attempt to explain such a high efficiency will be made.

An analysis of total adjusted plant height of all random selections from crosses with one parent in common was undertaken, using the following form:

| Source | df |
|---|---|
| Crosses | 8 |
| Selections | (s-1)-8 |
| Total | s-1 |

where s = number of selections involving one parent. Total adjusted heights were used rather than mean heights because the totals were readily available from the summary sheets made up by the Statistical Laboratory. Ten analyses, one for all the crosses of each parent, were completed for the plant height results of each year, the primary purpose being to determine whether more variation existed between the seven or eight selections from each cross involving one common parent or between the nine such crosses.

In 1951 there was significant differences between mean heights of crosses involving seven of the parents; this indicates that there was more variation between crosses than between random selections of each cross. In the case of the crosses involving Blackhull, Chiefkan, and Tenmarq there was no significant difference between crosses, indicating that as much variation existed within crosses as between crosses. In 1952 there was more variation between the crosses involving a common parent than there was within crosses in seven out of ten cases. The crosses involving Comanche, Pawnee, and Chiefkan had as much variation within as between crosses. In the 1952 test the LSD showed the plant height of the two Pawnee selections to be significantly different; otherwise there was no significant difference between the two selections of each parental variety in either year.

Since, in all cases, variation between crosses was greater than, or equal to, variation within crosses and in only one case was there a significant difference between selections from a variety it seems reasonable to judge the relative performance of a cross or variety by the average performance of all selections from that cross or variety. The best available estimate of the performance of each cross or variety should be obtained by averaging the mean height of all selections from each cross or variety over the two years in which selections were grown. When handled in this manner, the average height of the 55 kinds ranged from 43.0 to 38.1, a spread of 4.9 inches. The selections from entry 28, the Chiefkan parent, averaged tallest; those from entry 41, Pawnee, were shortest. Chiefkan had been among the tallest kinds in all early generation tests; Pawnee had been among the shortest.

Date of Flowering. Significant differences in date of flowering were found to exist in both years of selection testing. Adjusted date of flowering, with 1 equal to May 1, in 1951 ran from 34.6 to 20.6, a range of 14.0 days; in 1952 the range was 29.7-14.3 = 15.4 days. The five per cent LSD for comparing the difference between two mean dates of flowering was 2.05 days in 1951; 1.68 days in 1952. One of the selections from entry 51, Nebred x Marquillo-Oro, was latest in both years. Entry 51 had tended toward lateness during early generations.

The relative efficiency of the cubic lattice design compared to randomized complete blocks was 103 per cent in 1951 and 131 per cent in 1952.

Analyses of variance of total adjusted date of flowering of all the crosses involving one parent indicated, for 1951, that in the case of seven of the parents there was more variation between crosses than within crosses; in two cases, Early Blackhull-Tenmarq and Chiefkan crosses, the variation within was equal to the variation between crosses; and in the case of all the crosses involving Comanche the variability between selections from crosses was greater than the variability between crosses. Much the same situation existed in 1952 when all the crosses involving Early Blackhull-Tenmarq, Chiefkan, and Comanche were found to have variation between randomly selected lines equal to the variation between crosses. These same three parental varieties exhibited this high inter-line variation in both years. As in 1951, all the crosses involving the other seven varieties as common parents showed greater variation between than within crosses in 1952. In 1951 the LSD showed the two random selections from Comanche to differ significantly in date of flowering; no other case of significant differences between dates of flowering of selections from varieties was found in either year.

The variation between crosses involving one parent in common was greater than, or equal to, the variation between randomly selected lines of each cross in all but one case over the two-year period. Only once during the same period was there a significant difference between the two selections from any parent. On the basis of these facts it seems that a reasonable evaluation of the flowering date of any of the crosses or varieties would result from averaging the date of flowering of all selections from that cross or variety. The best available estimate of the performance of any cross or variety during 1951 and 1952 would be the two-year average of the mean performance of all selections from that cross or variety. When handled in this manner the average date of flowering of the 55 kinds ranged from 31.5 to 20.6, a spread of 10.9 days. The average date of flowering of the selections from Marquillo-Oro, a late parent, was the latest of any of the 55 kinds; the selections from Early Blackhull-Tenmarq, the earliest parent, averaged earliest over the two years. Marquillo-Oro, entry 53, was consistently the latest kind during early generations; Early Blackhull-Tenmarq, entry 55, was consistently the earliest during the same tests.

Test Weight. Significant differences between test weights of randomly selected lines were observed in 1951 and 1952. Adjusted test weights had a range of 59.7-51.8 = 7.9 pounds per bushel in 1951 and 65.6-57.8 = 7.8 pounds in 1952. The five per cent LSD for comparing the difference between two mean test weights was 1.03 and 2.87 pounds in 1951 and 1952, respectively.

For the test weight characteristic the cubic lattice design was 25 per cent more efficient than randomized blocks in 1951 and 39 per cent more efficient in 1952.

Analysis of adjusted test weight totals for all the crosses with one parent in common was completed as for plant height and date of flowering. In 1951 these analyses showed nine cases in which there was a significant difference between crosses and one case, the crosses with RedChief as the common parent, in which there was equal variation between and within crosses. Eight cases of greater variation between than within crosses were found in 1952; all the crosses involving Blackhull or Cheyenne as a common parent showed less variation between crosses than between selections from each cross. In 1951 the two selections from Chiefkan and the two selections from Marquillo-Oro differed significantly in test weight. There was no difference between the two random selections from each variety in 1952.

The average test weight of all selections from a cross or variety is suggested as a suitable measure of the performance of that cross or variety, since in all but two cases during the two years there was more, or equal, variation between crosses than out of twenty was there a significant difference between selections from a variety. The mean cross or variety test weight averaged over the two-year period should give the best available estimate of the performance of that cross or variety after selection. The range of test weights of the 55 kinds averaged in this manner ran from 61.6 to 56.2, a spread of 5.4 pounds per bushel. Selections from entry 35, the RedChief parent, had the highest average test weight of selections from any kind; selections showing the lowest average test weight were those from the Marquillo-Oro parent, entry 53. RedChief was the heaviest kind in practically all early generation tests; Marquillo-Oro was consistently the lightest kind during early generations.

### Early Generation Performance Compared to Performance of Selections

It has been proposed that the most practical way a plant breeder has to accumulate information on early generation bulk hybrids is to grow successive generations of the bulk hybrids in successive years. A combined analysis of F3 (1946), F4 (1947), and F5 (1948) generation results for each of four characteristics of 55 kinds has been included as a part of the early generation studies just reported. The studies of random selections from each bulked cross and parental variety have been reported. The best available single estimate of the performance of selections from each kind appeared to be the two-year average of the mean performance of all selections from each kind. With these two values available it becomes possible to evaluate the early generation bulk hybrid tests as predictors of the performance of selections from those bulks.

The product-moment correlation between the average yield of the 55 kinds during early generations and that of selections from each kind was calculated to be 0.2787, a significant correlation coefficient. It will be recalled that the analysis of variance of the early generations showed no significant difference between the yields of the 55 kinds when averaged over three successive generation-years. It would not have been possible, then, to classify any kind, or group of kinds, as high or low yielding; in the absence of any such preselection classification it would seem that the correlation between early generation yields and average yields of selections actually has little meaning for the purposes of this study.

The average successive early generation kind plant height was correlated with the two-year average of mean plant height of all selections from each kind; the correlation coefficient was calculated to be 0.7302, a highly significant value. This would indicate that random selections from each kind tended to have an average plant height similar, relatively, to that of the kind from which they were selected.

The product-moment correlation between kind dates of flowering averaged over three successive early generation-years and two-year average of mean dates of flowering of all selections from each kind was 0.8271, a highly significant value. The relative average date of flowering of random selections from each kind was similar to the relative date of flowering of these 55 kinds during their early generations.

Average test weight of grain from the 55 early generation kinds was correlated with the two-year average of mean test weights recorded for all selections from each kind; the correlation coefficient was calculated to be 0.8296, highly significant. It is indicated that the relative test weight of kinds averaged over three successive generation-years was similar to the relative mean test weight of random selections from each kind when averaged over the two-year period.

## DISCUSSION

Atkins and Murphy (2) and Weiss, Weber, and Kalton (23) successfully classified oat and soybean crosses according to yield during early generations, but found such classifications to be of no value for predicting the yield of selections from these crosses. Bjaanes (3); Harlan, Martini, and Stevens (9); and Harrington (11) reported successful yield classification of wheat and barley crosses based on early generation tests and found such preselection classifications to be of value as indicators of the yield of selections from these crosses.

It was not possible, on the basis of early generation tests of the 55 kinds studied in this experiment, to classify any kind, or group of kinds, as high or low yielding. Generation to generation and year to year inconsistencies in relative yield were too great to allow one to rely on any single test or any logical combination of tests as indicators of yield during early generations. Such yield inconsistencies are no doubt due to the interaction of many factors, both environmental and genetic, that operate to influence yield. Conditions encountered by growing wheat plants in any two years are never the same; yield response is apparently influenced, both directly and indirectly, by these conditions.

The apparent lack of year to year stability in relative yield of wheat varieties is well illustrated by Salmon (17), who found highly significant year x variety interactions in nine experiments conducted at five locations in the Great Plains over periods of ten to twenty-six years. Year x variety interaction for all possible pairs of varieties ranged from non-significant to highly significant, with a preponderance of the latter. Differences in certain factors, such as disease and insect resistance, maturity, and winterhardiness, are advanced to help explain year x variety yield interactions, but it is doubtful whether known factors entirely explain these interactions. Salmon (17) points out that, because of year x variety interactions, agronomists and plant breeders interested in the release of improved wheat varieties cannot practically test varieties over long enough periods to establish statistically significant differences in yield alone. Generally the quality of grain, other agronomic characteristics, and reaction to factors indirectly influencing yield are considered, along with available yield information, before a variety is released.

Since preselection classification of the 45 bulk hybrids and ten parental varieties according to yield could not be accomplished, it was concluded that the early generation bulk hybrid tests were of no value in predicting the yield that might be expected from selections from the 55 kinds. The probability of obtaint a high yielding line by selection was apparently equal, no matter from which cross or variety the selection might have been made.

Preselection classification of the 55 kinds according to plant height was found to be possible. Correlations show determinations of plant height made during any one generation or year to be reliable indicators of the relative plant height of the same kinds grown in other generations or years; it might be argued that a single test is sufficient to determine early generation performance. A better evaluation of early generation performance was probably obtained when the average plant height of the 55 kinds, grown in three successive generations during as many years, was employed. Such an average, covering a random sample of years, took into account the year x kind interaction shown to be present.

Within the limits of this experiment it has been shown that preselection determinations made on early generation bulk crosses and varieties reliably predict the relative plant height of selections from those crosses and varieties. If the material entered in this experiment had been included in a plant breeding program with the isolation of segregates with short straw as an objective, it would have been possible to increase the probability of success by discarding the taller early generation kinds before making any selection. The results in regard to plant height agree with those of Weiss, Weber, and Kalton (23) who found tests of bulk populations of soybean crosses to be indicative of the performance of selections from those crosses.

The 55 kinds could be classified into maturity groups on the basis of early generation date of flowering results. Correlations showed that the date of flowering during any generation or year was a reliable indication of the relative date of flowering that could have been expected when the same kinds were grown in other generations or years. A better estimate of the kind date of flowering during early generations probably comes from averaging kind dates of flowering over three successive generation-years. The year x kind interaction, believed to be caused by a few kinds, is partially accounted for in such an average.

It has been shown that the preselection classification of 55 kinds according to date of flowering reliably predicted the performance of random selections from those kinds. With the material studied, a plant breeder could have discarded all the late crosses from his program and, as a result, increased his chances of success in selection, if earliness had been his objective.

Results of the date of flowering study are apparently in line with the experience of Ackerman and MacKey (1). Weiss, Weber, and Kalton (23) report no success in predicting maturity of soybean selections from the results of bulk population tests; they explain, however, that the actual maturity of soybeans in their experiment was frequently not fully expressed because of the occurrence of frost.

Available data indicates that preselection classification of the 55 kinds according to test weight of grain was possible. Correlations show that test weight determinations during one generation or year are reliable as indicators of relative test weight of the same kinds in other generations and years. Much of the year x kind interaction is believed to have been accounted for by using the average test weight of three successive generation-years as an indication of early generation performance.

The preselection classification of 55 kinds according to test weight reliably predicted the performance of random selections from those kinds. The probability of selecting a high test weight line from this group of material would have been increased if low test weight kinds had been discarded at the end of early generation testing. Atkins and Murphy (2) found test weight of oat selections to be reliably predicted by early generation test weight of bulked crosses.

Other workers conducting studies similar to the one reported herein have attempted to determine the relationship between the characteristics studied. No such attempt has been made in this experiment; each characteristic has been studied independently. It is conceivable that some association does exist between certain of the characteristics studied, but it is doubtful whether there is close enough relationship to allow selection for one of the characteristics by selection based on another characteristic. Agronomists generally agree that the probability of success in selection decreases as increasing numbers of characteristics are added as objectives in a breeding program; it is anticipated that a similar conclusion would have been made in this experiment.

The results reported for this experiment suggest no drastic changes that should be made in present methods of handling segregating generations of winter wheat crosses; rather, they seem to support continuation of the present methods. While plant height, date of flowering, and test weight results were accurately predicted by bulk hybrid tests it is doubtful whether a practical plant breeder would find the increased accuracy worth the extra measurements, records, and analyses involved. It was shown that these characteristics are fairly consistent from year to year and it seems, therefore, that visual selection for plant height and date of flowering during segregating generations handled by the line method of breeding would have been more economical and fully as effective in securing the desired segregate. A single indicative test weight measurement could be made on the small amount of seed produced during one of the segregating generations of crosses handled by the line method; this experiment shows that such a determination would be fairly reliable as an indicator of test weight to be expected in later generations and years. It is suggested, however, that satisfactory test weight is likely to be one of the attributes of a segregate remaining after continued selection has been made for desirable plant type and agronomic characteristics.

Yield tests during early bulk generations were found to be of no predictive value. It is suggested that selection during segregating generations be based on characteristics other than yield; and that a series of yield tests be conducted, if necessary, to choose between pure lines that have already been selected for such characteristics as disease resistance, insect resistance, plant height, maturity, strength of straw and quality of grain. It is apparently feasible to make progress in selection for the aforementioned characteristics and it is likely that satisfactory yield will be manifest by most of the remaining segregates after selection for other observable and measurable characteristics has been accomplished.

Careful choice of parents to be used for making a relatively few crosses, followed by selection of desirable segregates during each early generation, is undoubtedly more economical and probably attains the desired result as quickly as a program involving the study of many crosses, some of which are discarded on the basis of their bulk performance. This report, concerning the latter method, indicates that visual selection for plant height and date of flowering will be effective in the former method.

It is possible, from the results of the experiment reported herein, to gain some insight as to the heritability of the four characteristics of winter wheat studied. Heritability might be defined as "the ability of a given characteristic to be transmitted from parents to the progeny." Low heritability of a characteristic is generally associated with a large number of genes and/or considerable environmental influence affecting that characteristic. Increasing heritability may be the result of fewer factors conditioning a characteristic and/or less response of that characteristic to environmental influence. Low heritability of yield is evident in this study; yield seems to be conditioned, both directly and indirectly, by a great many factors and is influenced a good deal by environment. Plant height, date of flowering, and test weight of grain in winter wheat appear to be highly heritable. Each of these characteristics may be conditioned by fewer genes; at any rate there seemed to be rather small environmental influence on the transmittal of them from parent to progeny under the environmental conditions encountered in this experiment.

## SUMMARY

Early generation tests on 55 kinds of winter wheat, including 45 bulked crosses and their ten diverse-type parental varieties, were conducted in 1946, 1947, and 1948 at Manhattan, Kansas. F3 generation bulks were tested in all three years, F4 generation bulks in 1947 and 1948, and F5 generation bulks only in 1948.

Seven or eight selections, made at random from each space-planted F5 generation bulk cross, were tested, along with two random selections from each variety, in 7x7x7 cubic lattice designs in 1951 and 1952.

Results of the early generation trials were compared with those of selection tests to determine whether or not the early generation tests had indicated the performance that might be expected from selections from each cross. Detailed analyses and comparisons of yield, plant height, date of flowering, and test weight are reported.

Yield differences between the 55 early generation kinds were found to be inconsistent from generation to generation and from year to year. It was not possible to reliably designate any kind or group of kinds as high or low yielding because of high year x kind interaction. A combined analysis of three successive generations grown in as many years showed no significant yield differences between the 55 kinds; therefore, it was not possible to classify these 55 kinds according to yield during early generations. In the absence of such a classification it was concluded that early generation yield determinations had been of no value in predicting the performance of pure line selections from the 55 bulked kinds.

Consistent differences between plant heights, dates of flowering, and test weights of the 55 early generation kinds were found to exist; it was considered feasible to classify these kinds according to each characteristic. Preselection plant height of kinds was related to the mean plant height of all selections from each kind, as shown by the correlation coefficient of 0.7302. Similar comparisons showed correlation coefficients of 0.8271 for date of flowering and 0.8296 for test weight. It was concluded that preselection classification of these 55 kinds according to plant height, date of flowering, or test weight was a reliable indication of the mean performance, for that characteristic, to be expected from all selections from each cross.

On the basis of this study, it seems questionable whether the use of a plant breeding system in which many crosses are made and some later eliminated on the basis of early generation bulk performance is as economical in securing a desirable segregate as other, more commonly used, plant breeding methods.

## ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to those persons who made the completion of this study possible. Advice, encouragement, and constructive criticism on this manuscript and the work it represents has been offered by Dr. E. G. Heyne, major instructor, who has inspired the author to attempt to continue work along scientific lines. The cubic lattice experiments used in this study were designed by Dr. H. C. Fryer, station statistican at Kansas State College. Cubic lattice analyses made at the Statistical Laboratory, Kansas State College, were under the supervision of Professor Henry Tucker. Both men gave freely of their time and advice on statistical matters during the course of this study. The author's wife, Betty, was a constant source of encouragement; her aid in checking the calculations, as well as in typing and correcting the original draft and tentative copies of this paper, was invaluable.

A final word of appreciation is directed to the Kansas Crop Improvement Association, which made funds available for the assistantship under which the author completed the work reported herein.

## LITERATURE CITED


1. Ackerman, A., and J. MacKey.
   The breeding of self-fertilized plants by crossing. Svalof 1886-1946, p. 46-72. Edited by A. Ackerman, O. Tedin, and K. Froier. Lund, Sweden: Carl Bloms Boktryckeri, 1948.

2. Atkins, R. E., and H. C. Murphy.
   Evaluation of yield potentialities of oat crosses from bulk hybrid tests. Agron. Jour. 41:41-45, Jan., 1949.

3. Bjaanes, M.
   Studies in spring wheat breeding. 42nd Report, State Expt. Sta., Moystad, Norway. Oslo: Grondahl and Sons Boktrykkeri, 1951.

4. Cochran, William G.
   Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 3:22-38, March, 1947.

5. ____________ and Gertrude M. Cox.
   Experimental designs. New York: Wiley and Sons, 1950.

6. Eisenhart, C.
   The assumptions underlying the analysis of variance. Biometrics 3:1-21, March, 1947.

7. Florell, V. H.
   Bulked-population method of handling cereal hybrids. Jour. Amer. Soc. Agron. 21:718-724, July, 1929.

8. Grafius, J. E., W. L. Nelson, and V. A. Dirks.
   The heritability of yield in barley as measured by early generation bulked progenies. Agron. Jour. 44: 253-257, May, 1952.

9. Harlan, H. V., M. L. Martini, and Harland A. Stevens.
   A study of methods in barley breeding. U.S.D.A. Tech. Bul. 720. Washington: Govt. Printing Office, 1940.

10. Harrington, J. B.
    Predicting the value of a cross from an F2 analysis. Canadian Jour. Res. 6:21-37, Jan., 1932.

11. ______________
    Yielding capacity of wheat crosses as indicated by bulk hybrid test. Canadian Jour. Res. 18(C): 578-584, Nov., 1940.

12. Hayes, Herbert Kendall, and Forrest Rhinehart Immer.
Methods of plant breeding. New York: McGraw-Hill, 1942.

13. Immer, Forrest Rhinehart.
Relation between yielding ability and homozygosis in barley crosses. Jour. Amer. Soc. Agron. 33:200-206, March, 1941.

14. Kalton, Robert R.
Breeding behavior at successive generations following hybridization in soybeans. Iowa Agr. Expt. Sta. Res. Bul. 358. 1948.

15. Love, H. H.
A program for selecting and testing small grains in successive generations following hybridization. Jour. Amer. Soc. Agron. 19:705-712, Aug., 1927.

16. Mahmud, Imam, and H. H. Kramer.
Segregation for yield, height and maturity following a soybean cross. Agron. Jour. 43:605-609, Dec., 1951.

17. Salmon, S. C.
Analysis of variance and long-time variety tests of wheat. Agron. Jour. 43:562-570, Nov., 1951.

18. Snedecor, George W.
Statistical methods. 4th ed. Ames: Iowa State College Press, 1950.

19. Sprague, G. F.
Early testing of inbred lines of corn. Jour. Amer. Soc. Agron. 38:108-117, Feb., 1946.

20. Suneson, C. A., and G. A. Wiebe.
Survival of barley and wheat varieties in mixtures. Jour. Amer. Soc. Agron. 34:1052-1056, Nov., 1942.

21. Tukey, John W.
Comparing individual means in the analysis of variance. Biometrics 5:99-114, June, 1949.

22. Weiss, Martin G.
Soybeans. Adv. in Agron. 1:78-158. 1949.

23. _______________, C. R. Weber, and Robert R. Kalton.
Early generation testing in soybeans. Jour. Amer. Soc. Agron. 39:791-811, Sept., 1947.

24. Yates, F., and William G. Cochran.
The analysis of groups of experiments. Jour. Agr. Sci. 28:556-580, Oct., 1938.

EVALUATION OF BULK HYBRID TESTS FOR PREDICTING PERFORMANCE OF PURE LINE SELECTIONS IN WINTER WHEAT

by

WAYNE LOVELLE FOWLER

B. S., Kansas State College of Agriculture and Applied Science, 1951

---

AN ABSTRACT
OF A THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Agronomy

KANSAS STATE COLLEGE
OF AGRICULTURE AND APPLIED SCIENCE

1953

Any increase in the efficiency with which segregating generations of small grain crosses are handled will contribute to the progress plant breeders are able to make in improving the small grains by hybridization. A testing system that would make possible the early identification and elimination of crosses with little or no potential value would allow more time and effort to be spent on the remaining, potentially valuable, crosses. It has been suggested that selection from bulked populations found to be desirable during early generation tests is more likely to yield a desirable segregate than selection from less desirable bulked populations.

Some workers have reported success of early generation testing for certain characteristics, including yield, in the small grains; others have found such tests to be of no value. It is the purpose of this paper to report the results of an experiment designed to determine the value of early generation bulk hybrid tests for predicting the performance of pure line selections in winter wheat. The experimental method involves a comparison between results of F3, F4, and F5 generation tests of the bulked F2 progeny of 45 winter wheat crosses and the results of two year's testing of random selections from those crosses. Detailed analyses are reported for yield, plant height, date of flowering, and test weight of grain.

Ten diverse-type winter wheat varieties were crossed in all possible ways in 1942. All the seed from F1 plants of each cross was bulked and increased for testing. The F3 generation bulks were grown in 1946, 1947, and 1948; the F4 generation bulks in

1947 and 1948; and the F5 generation bulk only in 1948. All early generation tests, which included the ten parents as well as the 45 crosses, were conducted in randomized complete block designs with five or ten replications. In 1949 ten selections were made at random from space planted F5 plants of each cross; two random selections were made from each variety. In 1951 and 1952 seven or eight of the selections from each cross were tested, along with both varietal selections, in a 7x7x7 cubic lattic design. All crossing, selecting, and testing was conducted at Manhattan, Kansas.

Highly significant differences between the 55 kinds were found for each of the four characteristics in all tests. Correlations showed that relative yield was inconsistent from generation to generation and year to year. The other three characteristics appeared to be fairly stable and a single determination probably would have been a reliable indication of the relative plant height, date of flowering, or test weight to be expected from the same 55 kinds grown in some other generation or year. The data most likely to be accumulated and used in a practical breeding program is that for successive generations grown in successive years. Accordingly, a combined analysis of F3 (1946), F4 (1947), and F5 (1948) generation results for each of the characteristics was run. No significant difference between the yields of kinds was observed; kinds were significantly different for the other three characteristics. The average performance of any one kind during these three successive generation-years was used as a measure of the performance of that kind during early generations.

Analysis of each year's results of selection testing showed significant differences between selections for each characteristic. It was demonstrated that, in practically all cases, there was greater variation between the nine crosses with any one parent in common than between the seven or eight selections from each of the nine crosses; therefore, it seemed reasonable to use the two-year average of the mean performance of all selections from each cross or variety as a measure of the performance of selections from that cross or variety.

The yield correlation between early generation bulks and selections from them was significant; however, it had been shown that there was no significant difference between yields of the 55 kinds averaged over three successive generation-years. The yield correlation, then, would seem to have little meaning for the purposes of this study. Correlations of plant height, date of flowering, and test weight between early generations and selections were all highly significant.

It was concluded that it was not possible to classify the 55 early generation bulk winter wheat crosses and varieties according to yield because of high year to year variation in relative yield. In the absence of a reliable preselection classification there was no basis for attempting to predict yield of pure line selections. Classification of 55 bulked winter wheat crosses and varieties based on plant height, date of flowering, or test weight seemed to be reliable and consistent; these preselection classifications were apparently accurate in predicting the plant height, maturity, or test weight that might be expected in selections.

It is questionable whether a prediction of plant height, maturity, or test weight of winter wheat selections made from the results of early generation bulk tests is worth the extra records and analyses entailed. Visual selection for plant height, maturity, and desirable type, coupled with selection for disease and insect resistance, during each segregating generation of fewer crosses handled by the line method would probably be as effective and more economical.