How Green Is Green? Sampling and Perception in Assessing Green Seeds and Chlorophyll in Canola

James K. Daun* and Stephen Symons

Canadian Grain Commission, Grain Research Laboratory, Winnipeg, Manitoba R3C 3G8, Canada

ABSTRACT: Green seeds are used as a grading factor in estimating chlorophyll in canola and rapeseed in the Canadian and U.S.A. grading systems. This work examines the effect that sampling and perception have on the estimation of green seeds as well as the effect that sampling has on the determination of chlorophyll. Individual seed analysis indicated that in order to be considered as green, seeds needed to contain between 200 and 400 mg/kg chlorophyll. Variation due to binomial sampling played a predominant role in the error in determining the green seed levels in canola. Sampling of large numbers of seeds, as in the loading of export shipments, reduced the error. Binomial sampling also contributed to the error in chlorophyll determination even with sample sizes as large as 500. Differences in perception of green also were noted between individuals with coefficients of variation as high as 50% at the 1% green seed level. The combination of perception error and sampling error may result in samples of 1,000 seeds drawn from a mass with 2% green seeds having green seed counts ranging from 0.96 to 3.04%, 19 times out of 20.

Paper no. J9561 in *JAOCS 77,* 1209–1213 (November 2000).

KEY WORDS: Binomial perception, canola, chlorophyll, grading, rapeseed.

Green seeds are the most predominant damage factor encountered when grading Canadian canola and rapeseed, probably accounting for up to 90% of the damage encountered in most years. Green seeds are considered damage factors in canola because they are indicators of the chlorophyll content of the overall sample (1). Chlorophyll is undesirable in canola because it is extracted with the oil, giving it a dark color which is expensive to remove (2). Chlorophyll pigments remaining in processed oils have been associated with increased oxidative rancidity (3) and difficulties in hydrogenation (4).

The quantity of green seeds is estimated in canola by transferring 100 seeds to strips of masking tape using a plastic strip sampling device containing 100 indentations each approximately the size of a single seed (5). The seeds are then crushed with a roller, exposing the internal meat of the seed. Seeds that are distinctly green in color are counted as damaged. Maximum tolerances for green seeds are 2, 6, and 20% for grades 1, 2, and 3 Canada, respectively. Samples with more than 20% green seeds do not qualify for statutory grades and are assessed as Sample Account Damage (Green). The difference in price for seed may be discounted as much as $10 to $15 per ton for each of the statutory grades, and sample seed is often discounted 50%. The tolerances have been set so that oil prepared from the top grade of seed will have less chlorophyll than the Canadian industry standard of 30 mg/kg in the oil (6).

Recognizing that there may be different interpretations of green seed, the *Official Grain Grading Guide* (5) defines green seeds as being "distinctly green." Pale green or otherwise immature seeds are taken into account in the evaluation of natural color. This means that samples which have less than 2% distinctly green seeds but which have a large number of pale green seeds can be downgraded on the basis of "poor natural color."

Assessment of green seeds in the grading of canola is a subjective process. Different people perceive color differently (7,8), and there are likely to be individual differences in perception of what is a green seed. For the purpose of providing uniform quality export shipments from a bulk handling system, the counting of distinctly green canola seeds has worked well in most years (Fig. 1) due mostly to the effect of the combining of large numbers of shipments from different areas of the country.

The actual relationship between distinctly green seeds and chlorophyll in the seed has been shown to be relatively poor (1) with correlation coefficients in the order of 0.7 for samples within the statutory grades. This paper explores some of the sources of error both in green seed counting and in the estimation of chlorophyll. In particular, the effects of the binomial nature of the sampling, the sample size, and the effects of individual perception of what is a green seed are examined.

MATERIALS AND METHODS

Samples of canola used in this study were obtained from the commercial Canadian crop through the Canadian Grain Commission's harvest surveys of canola (9). Samples were received from Canadian producers, grain handling companies, crushing plants, and from the Canadian Grain Commission's (CGC) Industry Services Division's official cargo loading

*To whom correspondence should be addressed at Canadian Grain Commission, Grain Research Laboratory, 1404-303 Main St., Winnipeg, MB R3C 3G8, Canada. E-mail: jdaun@cgc.ca

Copyright © 2000 by AOCS Press

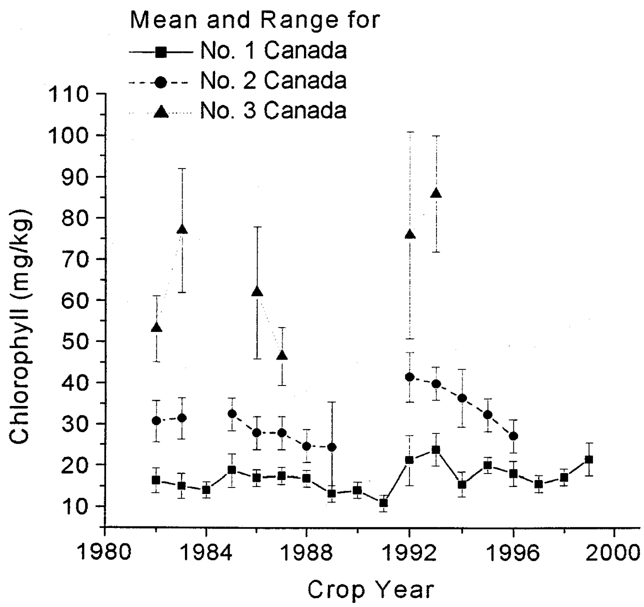

FIG. 1. Chlorophyll in Canadian export shipments of canola and rapeseed.

samples. Chlorophyll was estimated by the American Oil Chemists' Society (AOCS) Official Method Ak2-93 (10) or by analysis on a near infrared (NIR) spectrometer (11). These methods have been shown to give equivalent results. Green seed analysis was carried out as described by the CGC's *Official Grain Grading Guide* (5).

Analysis of the chlorophyll content of individual seeds was carried out by placing individual seeds into test tubes (10 × 75 mm), adding 1 mL of a mixture of heptane/ethanol (3:1), grinding with a Polytron homogenizer (Brinkmann Instruments Canada Ltd., Mississauga, Ontario), equilibrating for 1 h with gentle shaking followed by centrifugation and determination of the optical density at 700, 665, and 630 nm using a Spectronic 1001 (Spectronic Unicam, Rochester, NY) with a tube holder. The amount of chlorophyll was estimated by comparison with a standard curve made with pure chlorophyll a (Fluka, Sigma-Aldrich Canada Ltd., Oakville, Ontario). The method was subject to considerable error as the optical density readings were small for seeds with <50 mg/kg chlorophyll. Also, the small diameter of the test tube meant that slight differences in the positioning of the tube in the spectrophotometer resulted in significant variations in the optical density. The standard error of the method was estimated to be about ±10 mg/kg.

Statistical models. Binomial models were developed in Microsoft Excel using formula from Reference 12. Logarithmic conversion allowed large numbers to be manipulated within the restrictions of the program. Other statistical analysis was carried out using procedures of SAS v6.12 for Windows (The SAS Institute, Cary, NC).

RESULTS AND DISCUSSION

Effect of sampling error. The error of sampling for both green seed counting and for chlorophyll analysis is governed by the binomial nature of the sample. Seeds assessed in the grading of canola are either green or not green. This is a binomial type of experiment. These experiments have the following general properties: (i) There are n repeated trials (i.e., we will examine n seeds); (ii) each trial will result in a success or failure (the seeds will be green or not green); (iii) the probability of success remains constant from trial to trial. This is not strictly true in canola grading since each seed is not returned to the mass after assessment. CGC continuous sampling procedures ensure that samples obtained from rail car unloads or cargo loading are representative of the mass. Samples, laboratory samples, and test samples are all prepared according to methods which ensure that the eventual sample is random and representative. The ratio of experimental sample to the bulk is small, however, and it is possible to assume that the probability of success remains essentially constant. In addition, it is possible that the evaluation of which seed is green will not be consistent from inspector to inspector, or even for an individual inspector at different times; and (iv) the trials are independent. Again, this is not strictly true since the seeds are not returned to the bulk between assessments. However, the large volume of the bulk compared to the sample size means that the trials are essentially independent.

The binomial distribution states that the probability of finding a given number of green seeds (x) in repeated trials with sample sizes of n from a population where p is the proportion of green seeds and $q = 1 - p$ is

$$\Pr(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad [1]$$

The mean value of x is $\mu = np$, and the variance of x is $\sigma^2 = npq$. When the sample size is 500 seeds, as the minimum specified in the *Grain Grading Guide* for official CGC inspection, on the average 10 green seeds would be found for a 2% green seed count with a standard deviation of about ±3.1 seeds. For 95% of the time, we would expect 4 to 16 seeds or 0.8 to 3.2%.

One of the features of the binomial distribution is the discreteness of each cell. The distribution cannot be described as a continuous curve as can the normal distribution. The binomial distribution also is not symmetric but is skewed toward zero when $P < 0.5$ (Fig. 2). Likely, much of the variation of green seed counting in canola can be at least partially explained by the binomial type sampling employed. To verify this, at two different times, different seed lots were carefully subdivided into representative 100-g subsamples, using a Boerner divider (Hoffman Manufacturing, Inc., Albany, OR) and submitted to grain inspectors for an estimation of green seeds by the official grading method. In 1986, the head of the Grain Appeal Tribunal in Winnipeg analyzed twenty 100-g subsamples prepared from each of two different samples. In 1993, a grain inspector in the CGC's Vancouver office analyzed 32 different subsamples prepared from each of three different samples. In each case, the green seed count was from a 1,000-seed aliquot of each sample. The inspectors were not told that the subsamples were from the same lot. A chi-squared

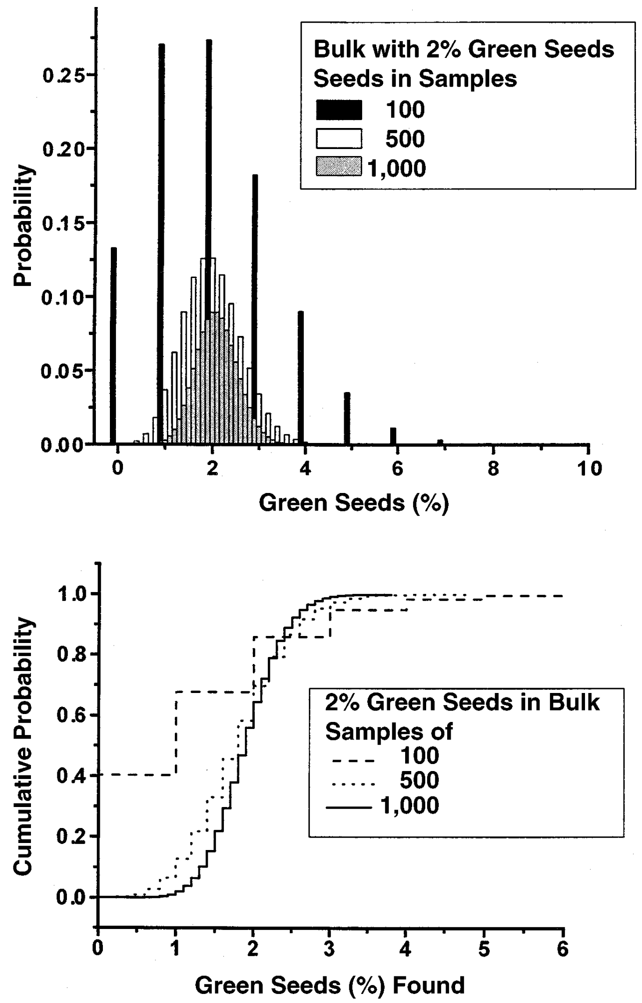

FIG. 2. Probabilities for green seeds found in different-size samples drawn from a bulk with 2% green seeds.

TABLE 1
Test for Goodness of Fit to the Binomial Distribution for Canola Green Seed Counting

	Sample[a]				
	V1	V2	V3	W1	W2
	Number of trials				
	32	32	32	20	20
	Green seeds				
Mean	1.5	3.2	5.5	1.0	1.6
Min.	0.6	2.4	4.4	0.0	0.7
Max.	2.4	4.4	7.1	1.9	2.4
df	14	18	19	10	10
χ^2 [b]	18.0	11.9	19.2	7.0	6.5
$\chi^2_{0.05}$	23.685	28.869	30.144	18.307	18.307

[a]V refers to samples counted in Vancouver in 1993, with V1, V2, and V3 representing individual replicates, and W to samples counted in Winnipeg in 1986, with W1 and W2 representing individual replicates.
[b]Goodness of fit test as described in Reference 12, pp. 252–256. $\chi^2 = \sum_{i=1}^{k}[(o_i - e_1)/e_i]$ where χ^2 is a value of the random variable X^2, whose sampling description is approximated very closely by the chi-square distribution. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the *i*th cell. df, degrees of freedom.

goodness of fit test showed that the distribution of results was not significantly different from the binomial (Table 1).

The dependence of the variability of green seed counting on the binomial distribution has some consequences for grading of canola. For counts from samples of 500 or 1,000 seeds, the skew of the distribution is minimized. For smaller counts (e.g., 100), however, there is a significant skew. Often, country elevator operators have been observed making grading assessments on as few as 100 seeds. This practice may work in favor of farmers making deliveries to country elevators as they will meet the 2% or less green seed requirement with a greater proportion of their deliveries than if a larger sample size were used. In samples drawn from a bulk with 2% green seeds, 2% or less would be found 67% of the time for 100 seed samples, 58% of the time for 500 seed samples, and 55% of the time for 1,000 seed samples (Fig. 2). In addition, the larger sample size gives a much smaller standard deviation. The situation also works for samples with true green seed values >2%. For example, samples with 4% green in the bulk would be counted as No. 1 Canada 23% of the time if the elevator operator counted 100 seeds, <10% of the time if 500 seeds were sampled, and <1% of the time if 1,000 seeds were sampled.

It is important that export shipments of canola be graded accurately. The effects of the binomial nature of sampling are reduced by the number of the samples analyzed during loading (about 1 sample for each 1,000 tons). For example, for an export cargo of 20,000 tons of seed, a total of about 20 subsamples each with 1,000 seeds counted from the continuous sampling would be evaluated for green seeds. If this cargo were found to contain 2% green seeds, then the 95% confidence interval for the final estimate would be a range of 1.8 to 2.2% due to the effects of sampling alone.

Because green seeds contain much more chlorophyll than nongreen seeds, the sampling distribution might be expected to affect the precision of methods for the analysis of chlorophyll in canola samples by the AOCS method.

Analysis of 500 single canola seeds from a bulk sample which was reported to contain 2% green seeds showed one seed having 420 mg/kg chlorophyll and five seeds having between 250 and 270 mg/kg (Fig. 3). To be considered a green seed, the seed would likely have to contain at least 200 mg/kg of chlorophyll and more likely as much as 400 mg/kg.

By using the levels of 200 mg/kg (minimum) and 400 mg/kg (maximum), it is possible to calculate the contribution of the binomial sampling of green seeds to the error in the determination of chlorophyll by the AOCS method. While the AOCS method specifies grinding of 50 g seed (*ca.* 15,000 seeds), it is more common to grind about 11.5 g seeds (*ca.* 3,500 seeds), or in the case of the original method (13), as little as 2 g (*ca.* 600 seeds). Even with a sample size of 50 g, the contribution of the binomial sampling to the error of the chlorophyll analysis would be significant, falling somewhere between 0.2 and 0.4 mg/kg at a 20 mg/kg level (Table 2).

The overall effect of sample size was confirmed by repeated analysis of samples of different size drawn from the

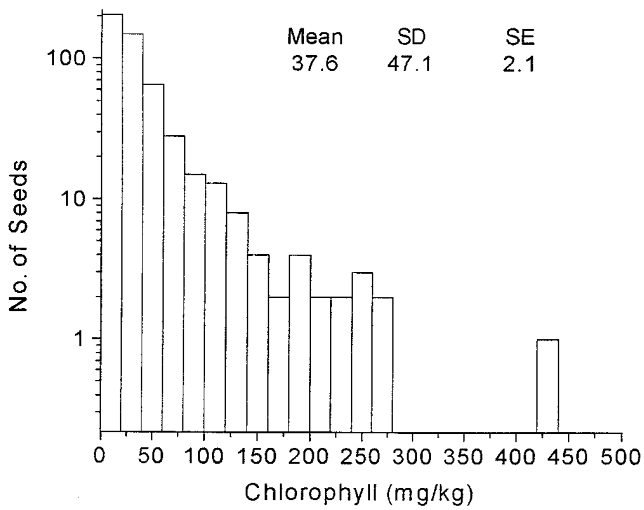

FIG. 3. Distribution of chlorophyll in a sample of 500 seeds from a lot with 2% green seeds.

same seed lots. The standard deviation increased with decreasing sample size (Table 3) as would be expected from effect of binomial sampling.

Effect of perception error. One thousand seeds were drawn from each of 12 samples of canola seed comprising three samples each of No. 1 Canada, No. 2 Canada, No. 3 Canada, and Sample Canola Account Damage. The green seed count for each sample was estimated independently by six CGC grain inspectors using the same set of strips crushed according to the method described in the *Grain Grading Guide*. A significant difference (analysis of variance) in the perception of distinctly green seeds was noted among inspectors (Table 4). Variance (Table 5) increased with the level of green seeds. Possibly providing inspectors with a color chart or crayon to assist in determining green seeds could reduce the error in perception; however, this would not remove the error due to the binomial nature of the sampling.

When the error due to perception between inspectors is coupled with the variation due to the binomial nature of the sampling procedure, the total error in green seed becomes significant. For example, binomial error for 1,000 seeds at 2% green is ±0.3% and the error due to perception, calculated from the mean variance in Table 4, is about ±0.42% for a total error of ±0.52% green. This means that 19 times out of 20 samples with true values of 2% would fall in the range of 0.96 to 3.04%, discounting the skew of the binomial distribution.

Even though the binomial sampling distribution also contributes to the error in direct chlorophyll measurement, this procedure still provides a more precise estimate of the chlorophyll level than distinctly green seed counts, and hence pro-

TABLE 2
Calculated Effect of Binomial Sampling on the Analysis of Chlorophyll in Samples where Nongreen Seeds Have 20 mg/kg Chlorophyll

Sample			Chlorophyll (mg/kg)		
			Calculated binomial effect on chlorophyll analysis[a]		
Weight (g)	No. of seeds	In a green seed	Min.	Max.	SD
50	15,000	200	23.19	24.01	0.206
		400	26.73	28.47	0.434
11.5	3,500	200	22.75	24.45	0.426
		400	25.80	29.40	0.899
2	600	200	21.54	25.66	1.029
		400	23.26	31.94	2.172

[a]Expected range and SD for a single analysis based on binomial sampling statistics.

TABLE 3
Effect of Sample Size on the Precision of Chlorophyll Determination by the AOCS Method Ak 2-92 (ref. 10)

		Chlorophyll (mg/kg)			
Sample	Weight (g)	Mean[a]	SD	Min.	Max.
A	1	19.3	2.6	13.8	23.2
A	2	20.5	3.4	17.0	31.8
A	10	20.8	0.4	20.2	21.9
A	50	20.2	0.6	18.9	21.7
B	1	36.8	3.7	31.0	45.4
B	2	36.9	2.5	33.2	41.2
B	10	37.4	1.3	35.7	39.4
B	50	38.9	0.9	37.2	40.7

[a]Based on 18 grinds followed by AOCS chlorophyll estimation at each level of sampling.

TABLE 4
Variability of Green Seed Estimates for Six Inspectors Assessing 1,000 Seed Subsamples Independently

		Percentage distinctly green seeds				
Grade	Chlorophyll (mg/kg)	Mean[a]	Min.	Max.	Range	Variance
1CC	19	0.93	0.6	1.8	1.2	0.20
1CC	9	1.10	0.8	2.2	1.4	0.30
1CC	14	1.13	0.8	1.4	0.6	0.04
2CC	43	1.82	1.2	2.6	1.4	0.24
2CC	28	3.58	2.8	4.8	2	0.51
2CC	39	3.77	3.4	4.2	0.8	0.09
3CC	72	9.73	8.4	11	2.6	1.11
3CC	83	11.53	10.2	12.2	2	0.70
3CC	64	11.85	10.2	12.8	2.6	1.01
Sample	122	14.57	12.2	17.8	5.6	3.96
Sample	121	20.20	17.6	22	4.4	3.78
Sample	136	21.27	16.4	24.2	7.8	8.39

[a]Mean results over analyses by six inspectors.

TABLE 5
Analysis of Variance Table[a]

Source of variation	SS	df	MS	*F*	*P*-value
Grades	3511.574	3	1170.525	1003.6	<0.0001
Samples within grades	185.218	8	23.1522	19.85	<0.0001
Inspectors	37.54	5	7.507	6.436	<0.0001
Error	64.15	55	1.166		
Total	3798.48	71			

[a]Abbreviations: SS, sum of squares; MS, mean square. See Table 1 for other abbreviation.

vides a better estimate of the intrinsic value of the seed lots. In addition, direct chlorophyll analysis, using whole seed NIR analyzers, can be carried out much more rapidly than green seed assessment. If chlorophyll analysis were adopted, however, it would still be necessary to estimate other factors such as heated seeds by visual assessment of the crushed seeds (5) unless an instrumental calibration for this characteristic can also be developed.

An understanding of the effect of binomial sampling on the assessment of green seeds in the grading process is important to the Canadian grain industry. Farmers and grain elevator operators should understand the nature of sampling where a difference of green seed between 1.9 and 2.1% may mean the difference of $300 on a 20-ton delivery. It is also important that the CGC inspectors and the Grain Industry understand the inherent variability due to both sampling and perception, especially for setting tolerances where samples are submitted for reinspection.

A more general conclusion from this study might be to point out the importance of sampling from discrete sources (such as seeds) when dealing with substances that are not uniformly distributed throughout the mass. Other examples such as pesticide-treated seeds, the presence of fungal damage, the presence of bin-heated seeds, the presence of admixture (especially inconspicuous admixture such as mustard in canola), sprouted seeds, etc., will also be expected to give similar effects. These effects will occur on the error in the estimation of the particular factor and may contribute to the estimation of chemical factors associated with the substance in question.

ACKNOWLEDGMENTS

Kathleen Clear carried out chlorophyll testing, Peter Edwards, formerly Chair of the Grain Appeal Tribunal and the staff of the Winnipeg and Vancouver offices of the Canadian Grain Commission's Industry Services Division, assisted in green seed counting. Jeff Babb, formerly of the CGC and now at the University of Manitoba, assisted with the statistics. Paper No. 805 from the Grain Research Laboratory.

REFERENCES

1. Daun, J.K., The Relationship Between Rapeseed Chlorophyll, Rapeseed Oil Chlorophyll, and Percentage Green Seeds, *J. Am. Oil Chem. Soc. 59*:15–18 (1982).
2. Mag, T.K., Canola Oil Processing in Canada, *Ibid. 60*:380–384 (1983).
3. Tautorus, C.L., and N.H. Low, Chemical Aspects of Chlorophyll Breakdown Products and Their Relevance to Canola Oil Stability, *Ibid. 70*:843–847 (1993).
4. Abraham, V., and J.M. Deman, Hydrogenation of Canola Oil as Effected by Chlorophyll, *Ibid. 63*:1185–1188 (1986).
5. *Official Grain Grading Guide,* Canadian Grain Commission, Winnipeg, 1998.
6. Daun, J.K., and C.T. Thorsteinson, Determination of Chlorophyll Pigments in Crude and Degummed Canola Oils by HPLC and Spectrophotometry, *J. Am. Oil Chem. Soc. 66*:1124–1128 (1989).
7. Mabon, T.J., Color Measurement in Food, *Cereal Foods World 38*:21–25 (1993).
8. Polger, T.K.W., Visualization and Color: Perception, Misperception, and the Next Generation of User Interfaces, *Proceedings of the User Interface Symposium (McLean, VA: The MITRE Corporation)*, http://www.duke.edu/~twp2/HCIC/visualization.html: The MITRE Corporation, 1993.
9. DeClercq, D.R., and J.K. Daun, *Quality of Western Canadian Canola 1998,* Canadian Grain Commission, Grain Research Laboratory, Winnipeg, 1998.
10. *Official Methods and Recommended Practices of the American Oil Chemists' Society,* 5th edn., edited by D. Firestone, AOCS Press, Champaign, 1998, Method Ak 2-92.
11. Daun, J.K., K.M. Clear, and P.C. Willams, Comparison of Three Whole-Seed Near Infrared Analyzers for Measuring Quality Components of Canola Seed, *J. Am. Oil Chem. Soc. 71*: 1063–1068 (1994).
12. Walpole, R.E., *Introduction to Statistics,* The Macmillan Com-

pany, New York, 1968.
13. Appelqvist, L.-Å., and S.A. Johansson, Fettvaliteten hos svenskt oljvaxtfro. II. Bestamning av klorofyllhalt i raps-, rybs- och vistenapsfor, *Sveriges Utsadesforings 77*:415–431 (1967).

[Received March 16, 2000; accepted August 3, 2000]