Quick Fix

FUZZY DATA SETS

WILLIAM D. STANSFIELD

The picture on the cover of *The American Biology Teacher* for January 2007 (by Martin Shields)* could serve as an icon for teachers of genetics. Four phenotypic classes of seeds from four varieties of peas are shown: yellow, round; yellow, wrinkled; green, round; and green wrinkled. This photo reminded me of an article in my files from 1984 by Julie Ann Miller.

> *Some seeds are smooth, and some are clearly wrinkled, but what about those with only a few dimples? They are the key to the problem of Mendel's data, says Robert S. Root-Bernstein* (1983) *... [He] asked 50 undergraduate biology students to count purple or yellow, wrinkled or smooth, kernels of maize from genetic experiments. He found that 6 percent of the kernels were classified as "indeterminant"* [sic]. *He cites a 1911 study in which fifteen trained geneticists were asked to analyze 532 yellow or white, starchy or sweet, maize kernels. That study reported,* No two of the fifteen highly trained and competent observers agreed as to the distribution of these 532 kernels.

Mendel (1866) did not report any "indeterminate" phenotypes in his famous paper on his experiments with pea hybrids. What became of these ambiguous peas? In a research report today, we would expect to see a tally that included seeds of questionable classification. Mendel made scores of experiments, many with hundreds of individual pea plants. This must have involved a great amount of time. Mendel was a teacher with lessons to prepare, lectures and/or demonstrations to give, papers and tests to evaluate, administrative reports to complete, etc., all of which greatly reduced the time he could spend on his experiments. It is known that the monastery had a gardener who helped Mendel from time to time. Miller (1984) suggests that: *His gardener may have doctored the data to please the monk.* It is not known if Mendel had any help from his students. If Mendel made no allowance for indeterminate phenotypes, his helpers would be forced to make subjective decisions regarding the phenotype, none of which likely would be identical. Whether or not this data recording problem had any influence on the "goodness of fit" of his experimental data to expected values remains debatable. I believe that students should be introduced to "fuzzy (subjective) data sets" such as this before they encounter them in their biology laboratories or in other research papers.

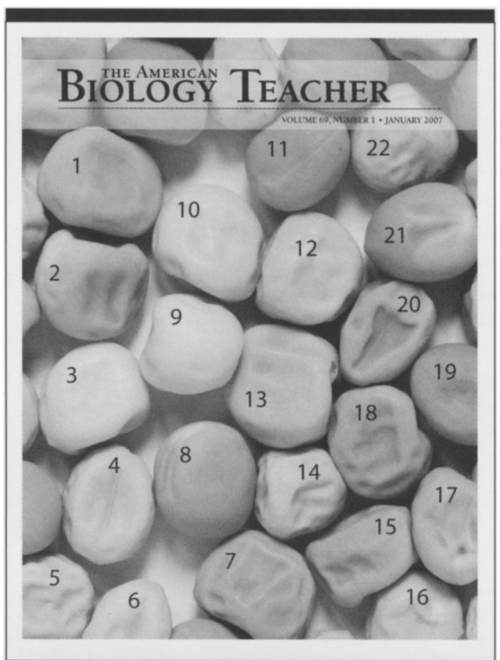


Figure 1. Copy of the cover photograph of *The American Biology Teacher* for January 2007. Shown are pea seeds numbered 1 through 22 from four varieties by the author.

I would like to suggest that teachers distribute copies of the ABT cover photograph (Figure 1) with 22 numbered peas, as part of a genetics laboratory or as a homework assignment. Each

STUDENT EVALUATIONS

Seed #	Phenotypic Classes				
	Yellow, Round	Yellow, Wrinkled	Green, Round	Green, Wrinkled	Indeterminate
14		X			
21			X		

Table 1. Example of a Student Evaluation form for Bobby Brown.

WILLIAM D. STANSFIELD, Ph.D., is Emeritus Professor in the Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA 93407; e-mail: wstansfi@calpoly.edu.

* ©Martin Shields. Readers may duplicate this image for use with their students. However, it may **not** be posted online. For inquiries, contact Photo Researchers Inc. at photoresearchers.com.

Downloaded from http://online.ucpress.edu/abt/article-pdf/69/8/458/54318/4452204.pdf by guest on 24 June 2021

student would try to classify each of the 22 seeds into four expected phenotypes; any phenotypes that students are unsure of would be tallied as "indeterminate." These data would be entered on a "student evaluation" form akin to the example in Table 1. Each student would then enter her or his data onto a "class summary" table akin to the example in Table 2. If this exercise is to be completed in a single class period, the class table could be on the blackboard. With large size classes, however, it might be more efficient to use two or more class summary sheets with students assigned into groups by their last names (e.g., students whose last names start with letters A through M enter their data on the sheet so marked; students N through Z on another sheet so marked). The master class summary could then be quickly compiled by the teacher and/or a designated aide. Copies of the master class summary could be given to each student at the next or subsequent meeting. Alternatively, the master class summary could be made into a transparency and displayed on a screen by an overhead projector.

CLASS SUMMARY

Seed #	Phenotypic Classes				
	Yellow, Round	Yellow, Wrinkled	Green, Round	Green, Wrinkled	Indeterminate
9	IIIII IIIII II				
19			IIIII III	III	I

Table 2. Example of a Class Summary sheet for a class of 12 students.

Once the students have access to the master class summary, the teacher can moderate a discussion session that highlights (among other topics) the problems of "fuzzy" data sets. For example, how "good" (reliable) is the data? Obviously, examining a picture is not as good as having the actual seeds that we can handle and thoroughly examine. We can only see about one-half of each seed in the picture. If we could see the whole seed, perhaps fewer of them would be classified as indeterminate. Admittedly, a very high resolution picture might have revealed more details. For example, Seed #17 on the cover picture appears to be more yellow than green near the hilum than elsewhere on that seed. Teachers may wish to make their own copies directly from the cover of *ABT*, perhaps with fewer selected peas numbered. If the cost of making enough colored copies for each student in the class is a problem, perhaps students could be asked to refrain from making any marks on their copies so that they can be reused for subsequent classes.

Here are some other questions for discussion. Are there any seeds whose phenotype was agreed to uniformly by every member of the class? Are there any seeds whose phenotype was scored as "indeterminate" by at least half of the students in the class? Should every seed that has at least one "indeterminate" score be eliminated from a paper for publication? If that was done, what percentage of the collected data would be discarded? Do all plants identified as "green" have the same intensity of color (hue)? Do the various hues suggest that a continuous spectrum of green colors (from light to dark) may exist in the population to which these seeds belong? If so, how can subjective data of this kind best be described? What possible factors could be responsible for this kind of variation? From the data alone, is it possible to identify either green or yellow as a dominant trait? How do you justify your answer?

References

Mendel, G. (1866). Experiments on Plant Hybridization. Bohemia: Proceedings of the Natural History Society.

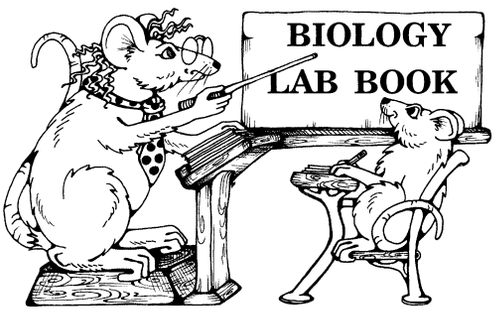
Miller, J.A. (1984). Mendel's peas: A matter of genius or of guile? *Science News*, *125*, 108-109.

Root-Bernstein, R.S. (1983). Mendel and mythology. *History of Science*, *21*, 275.

BRIDGE THE GAP

Advanced Placement Biology → College Credit

Biology Lab Book will work for you whether or not your school has a Concurrent Program with a local college.



- Used with great success in high schools offering college credit.
- Used by over 70 colleges & universities.
- All lab experiments work & adjust to your budget.
- Affordable lab books.

Check our website: **suspendedanimations.net.**
Or call us toll free: **1-877-468-4777** for more information.

Downloaded from http://online.ucpress.edu/abt/article-pdf/69/8/458/54318/4452204.pdf by guest on 24 June 2021