JOURNAL OF GEOSCIENCE EDUCATION **60**, 277–287 (2012)

ModeMaker and ModeQuiz: Tools for Enhancing Student Learning Estimation Skills of Rock-Component Abundance

David M. Hirsch[1,a]

ABSTRACT

Mode estimation, the visual determination of the proportions of the components of a rock, is a valuable tool in geology, providing one of the most efficient means of describing rock mineralogy and chemistry. However, current methods of teaching mode estimation suffer from the often-mediocre mode-estimation skills of the instructors. Fifty-nine academic and professional geologists was surveyed online to assess their mode estimation skills, and they scored 65.9 ± 7.5 (SD) out of a possible 100. Instructors with poor skills cannot provide the correct answers to mode questions and, thus, cannot create valid assessments of student mode-estimation skills. Two computer programs may help address this problem. ModeMaker is a tool for the instructor to use to create images with up to five phases, each with known modes. Each phase can have a number of additional specified properties, such as shape, size, and orientation. The instructor can, thus, provide valid assessments of student abilities to estimate modes. ModeQuiz is an interactive training application for the student. It creates images of the sort ModeMaker creates but randomizes the properties. The student makes mode estimations on the computer, and the program reveals the correct modes and scores the student's estimates. By repeating this process, the student may improve his or her mode-estimation skill over time. Assessments of ModeQuiz show that as little as 2 h of practice with the program may lead to improvement in student skill at mode estimation. © *2012 National Association of Geoscience Teachers*. [DOI: 10.5408/11-254.1]

Key words: mode estimation, software, skill learning

INTRODUCTION

When a geologist approaches and examines a rock, he or she typically asks three initial questions: What is the texture? What components are present? How much of each component is present? The first two of these are covered extensively in the typical undergraduate curriculum. The third—mode estimation—is the focus of this article.

To be sure, there exist quantitative methods of mode determination (*mode* is defined here, depending on context, as either the volume percentage of all the components of a rock, such as clasts and cements in a sandstone or minerals in a granite, or as the volume percentage of a single component: "mode of plagioclase"). The earliest effort in this arena was published by Delesse (1866) in which he described a method of tracing mineral outlines from a polished slab onto waxed paper and transferring the tracing to a sheet of thin metal, which would be cut along the boundaries, and the various piles of "minerals" weighed. *Point counting*, in which a set of regularly spaced points is examined and the identity of the component found at each point is recorded, has been widely used for decades, particularly when examining rocks under the microscope (Galehouse, 1971). Image processing may be used in certain situations where the phases can be distinguished based on color (Marschallinger, 1997) or backscattered electron signal (Dilks and Graham, 1985). Microbeam-based technologies can be powerful, offering phase composition in addition to modal abundance (Tinkham and Ghent, 2005).

Received 22 July 2011; accepted 5 March 2012; published online 13 August 2012.

[1]Geology Department, Western Washington University, Bellingham, Washington 98225, USA

[a]Author to whom correspondence should be addressed. Electronic mail: hirschd@geol.wwu.edu. Tel.: 360-650-2166. Fax: 360-650-7302

None of these methods is as efficient as the visual estimation of mode, which can be performed in seconds, requires no tools (save perhaps a hand lens), and can be easily performed in the field. However, such efficiency is of little value if the estimates are imprecise and/or inaccurate. For example, with poor mode estimation, one might erroneously infer that two granitoids were from the same batch of magma or that two sandstones had the same provenance.

This article explores mode estimation, assesses the skills of instructors and professionals and suggests possibilities for improvement in this aspect of the curriculum. In addition, it presents two new software tools that may aid in the teaching and learning of mode estimation and will assess one of those tools.

BACKGROUND

In many geoscience departments, students are taught the skill of mode estimation in a cursory fashion, or not at all. The instruction typically accompanies a laboratory focused on clastic sedimentary or intrusive igneous rocks, in which mode estimation is required for rock categorization (e.g., LeMaitre et al., 2005). In such laboratory exercises, the student is given a mode-estimation reference sheet (Fig. 1) and asked to estimate the mode for a set of rock specimens.

There are a number of weaknesses that may exist with this method. These include the potential lack of instructor knowledge of the true mode, a paucity of repetitions available to the student for learning, and the delay between the student's estimate and the feedback on that estimate.

Determining the actual mode of test or training specimens is potentially difficult. Even if an instructor is proficient in mode estimation, an error made in determining the true mode of a specimen used for student training may

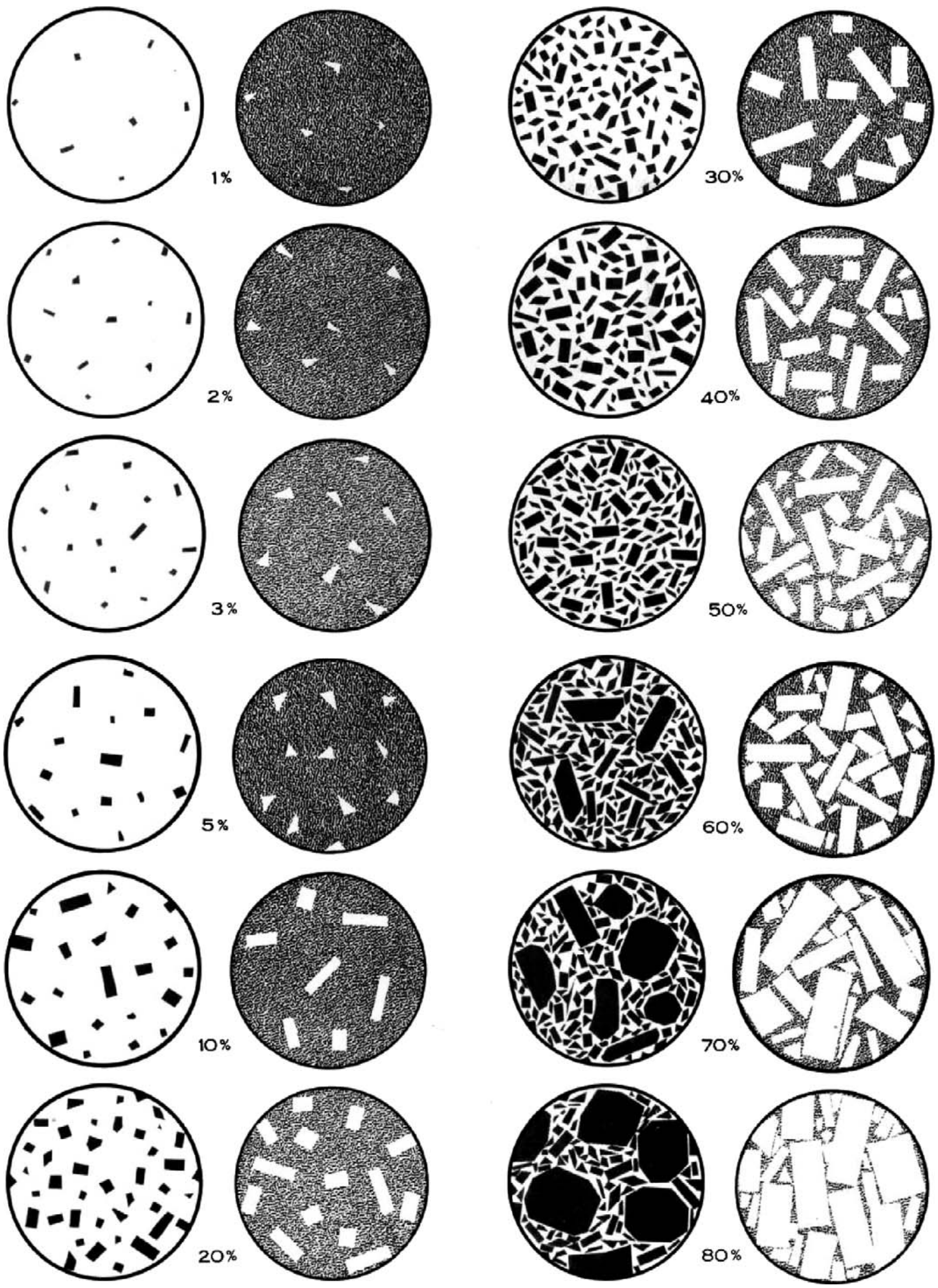


FIGURE 1: Sample mode-estimation reference sheet (Philpotts, 1989).

have far-reaching effects because many students may miscalibrate their mode-estimation sense. This weakness may be avoided at the expense of a substantial time investment by taking photos of specimens (or scans if they have a flat surface) and working with those photographs in image processing software to obtain quantitative measurements of the mode, if the components may be easily distinguished by color. For example, Adobe Photoshop (Adobe Systems Inc., San Jose, CA) and GIMP (The GNU Image Manipulation Program, http://www.gimp.org/) both have the ability to interactively create selections by color. By creating selections that correspond to each mineral and counting the pixels in each selection, the mode can be straightforwardly determined. However, creating such selections is not trivial, particularly with any shadows or reflections in an image, or when the color ranges corresponding to multiple minerals overlap. The author has used such image processing methods in the past to avoid his own weakness in mode estimation.

In addition, anecdotal data suggest that many instructors provide only a small number of opportunities for mode estimation. Unless an instructor believes himself or herself to be proficient at mode estimation, obtaining what is believed to be an accurate mode for a specimen can be laborious, as mentioned above. Many instructors will, thus, only provide students a small number of specimens for training themselves in mode estimation by making estimates and comparing them to the correct answer. Compounding these difficulties, after a student has made estimates for the mode of a rock and received feedback, the student can no longer use that rock for mode estimation, because he or she knows the answer.

As well, the time lag typical in a laboratory setting between the estimate and the feedback is not conducive to student learning. Pedagogical research has shown that delay in feedback can impair student learning when compared with rapid feedback (Kulik and Kulik, 1988). When the instruction in mode estimation is in the form of part of a laboratory exercise followed by feedback after the exercise is graded, the delay would appear to be unavoidable. Some laboratory exercises may be set up to provide immediate feedback on mode estimates, but, in the author's experience, this type of arrangement is rare.

To facilitate student learning of mode estimation and to avoid the weaknesses mentioned, it would appear that two needs must be met: (1) the ability to make a large number of rock-like images with well-characterized modes to work around the difficulty in obtaining true modes for actual rocks, and (2) the ability for students to obtain rapid feedback on their estimates for a large number of images. The first need is addressed by a new computer program, ModeMaker, the second, by a related but separate program, ModeQuiz.

MODEMAKER

ModeMaker is a computer program designed to allow instructors to create mode diagrams of the sort typically used for reference purposes (Fig. 1). It allows more-complex diagrams to be created as well. It is intended to allow the user to create images that broadly resemble rock textures, although it does not have the ability to create detailed textures. It is not (at this time) designed to be a crystallization simulation of any sort, and many real crystal shapes are not available (e.g., amphibole rhombi, pyroxene octalaterals). In addition, it cannot create textures such as reaction rims or depletion halos nor can it mirror metamorphic textures, such as the deflection of foliation paths around porphyroblasts.

Program Operation

The core of the software is the ability to create an image with up to five "phases," in addition to the background "phase," each occupying a known fraction of the image. The conceptual model is that of transparent layers on a canvas of user-specified size (Fig. 2). Each phase is independent of the others and occupies a separate layer. The user may alter the ordering of phases as desired (vertical arrows in Fig. 2).

Numerous properties of each phase may be specified. Some properties have fixed values (e.g., color, shape), whereas others have a mean and standard deviation (e.g., size, aspect ratio). For the latter, the value of the property assigned to any individual particle is a random selection from the normal distribution specified by the mean and standard deviation, typically with some restriction, such as that the property be positive.

Properties that govern the shape of each particle include the shape type, complexity, aspect ratio, and reentrancy. The particle shape may be an ellipse, rectangle, polygon, or "blob," which is a polygon with rounded vertices (Fig. 2). The number of points that make up a blob or polygon phase is termed *complexity* and is specified as the mean and standard deviation of a normal distribution, with the hard-coded requirement that the number of points be between 3 (to make a visible shape) and 12 (to prevent overloading of the CPU). Blobs and polygons may be required to be convex or may allow reentrants. The phase may have an *aspect ratio*, taken as the ratio of the long dimension to the short dimension (i.e., greater than or equal to unity). The aspect ratio is given as a normal distribution with the requirement that it be at least unity. Circles and squares may be created by fixing the aspect ratio to unity with zero deviation.

A "fabric" may be imposed on a phase if it has an aspect ratio greater than unity. This is a value from zero to one, with zero being no fabric at all (the long dimension of the particles are randomly disposed) and one being a perfect fabric (the long dimensions of the particles are all horizontal). Only horizontal fabrics may be imposed.

The principal property is, perhaps, size, which is specified as a characteristic linear dimension and is also a normal distribution, with the restriction that particle sizes be positive. For rectangles, this size value is the square root of the area; for other shapes, this size value is the square root of the area divided by $\pi/4$, which would cause the size value to be the diameter were the shape a circle.

The mode of a phase is not a property of each particle but a target for particle production; particles are produced until the mode requirement is satisfied. The mode is specified with an error value; that value indicates the precision with which the target mode must be achieved. A factor complicating the mode targeting is the manner in which the mode is measured by the program (see discussion below). Related to mode is the ability to specify whether individual particles may overlap each other. Preventing overlap causes particles to be separated from each other, if only by a small distance.

Once the user has created a suitable image, he or she will typically wish to export it to some other environment, such as a presentation or word-processing program. Export may be to a bitmap graphic format, the Portable Network Graphics (PNG) format, or to a vector graphic format, the Portable Document Format (PDF). If another format is desired, numerous free, third-party graphics converters are available for all platforms (e.g., GIMP at http://www.gimp.org).

The program has been designed for usability by people familiar with commercial software. Important functions have keyboard equivalents for efficient use. All controls have help tags, or "tool tips": short informative text about the function of the control that appears when the cursor hovers over the control briefly. Finally, the program includes internal "help" pages available from the Help menu. These pages serve as the documentation for the program as well, and include this manuscript among other information, such as the release history, extended notes on certain aspects of program operation, and plans for future improvements.

Relevant Program Internals

Although most aspects of the internal operation of the program are not sufficiently relevant to the user to merit

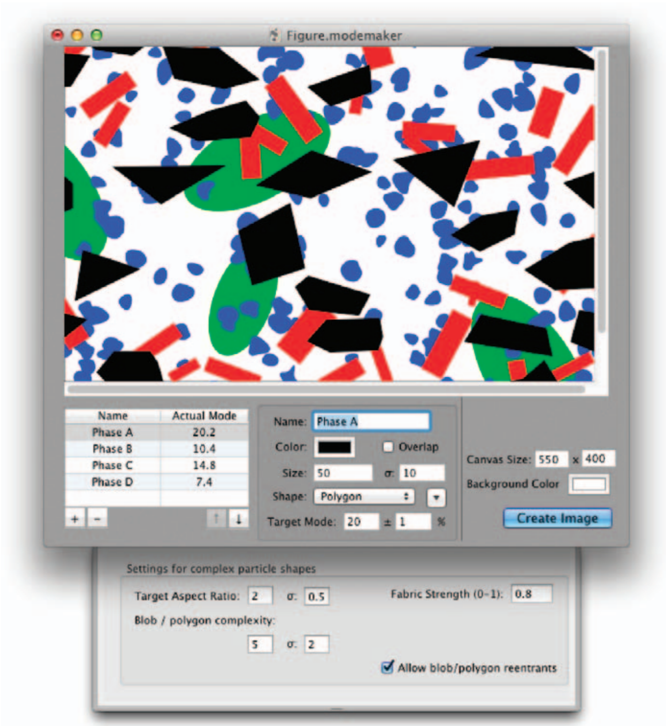


FIGURE 2: Sample ModeMaker window, showing four phases documenting some of the range of settings available.

inclusion here, some must be mentioned because their details affect parts of the user experience. These aspects include the manner in which modes are measured, the way in which edge effects are avoided, the conditions under which particles are recreated, and how blobs' shapes are created.

The most important part of the program, of course, is the measurement of the fraction of the image occupied by each phase. Under simple conditions (a single phase without overlap), direct calculation of the image fraction would be possible. However, as soon as particles can interact with each other, either by overlap between particles in a phase, or the obscuring of particles in one phase by those of another in a "higher" layer, the calculation becomes prohibitively complex and time-consuming. In lieu of such a calculation, therefore, the measurement of the mode is accomplished by a Monte Carlo method. In this method, a large number of randomly located points within the canvas are created, and for each, it is determined which phase, if any, occupies that point. Thus, the measurement of mode is imperfect, and carries some uncertainty. There exists a trade-off between the speed of program operation and the accuracy of the mode measurement. During program construction, exploration revealed a satisfactory compromise between these goals to be reached at 10,000 points per measurement. This provides a margin of error of ±0.69% for image fractions near 0.5, decreasing to ±0.14% for image fractions near 0.01 or 0.99, all with 95% confidence. With 95% confidence, the actual mode of a phase can thus be as far off the target value as the sum of the mode target error specified by the user (mentioned above) and the margin of error.

Were the particle centers to be placed randomly only within the visible canvas area, the canvas edge would look different from the interior, and mode estimation would be made more difficult. The limitation of having particle centers only within the canvas would lead to a lower-than-average image fraction near the edges because those pixels near the edge would be unlikely to be covered by particles located edgeward from the pixel. In addition, to compensate for this dearth near the edges, the image fraction in the interior would necessarily be greater than the target mode to achieve the desired target mode for the canvas as a whole. This would produce a gradient in image fraction from the canvas center to the edge, making mode estimation substantially more challenging. To avoid these problems, particle centers are placed on a "super-canvas" that is 10% larger in both dimensions. This allows particles to appear to graze the canvas edge from the outside, making the mode constant across the canvas.

For many sets of specified parameters, calculating or recalculating a phase is sufficiently costly, computationally, that it takes appreciable user time during which an "in-progress" cursor might be displayed by the operating system. ModeMaker thus attempts to minimize such recalculation, by only performing the recalculation when the canvas is resized or when the phase itself or a higher phase has been altered in a way that requires recalculation (e.g., changing the color will not require recalculation, but most other changes will). For example, suppose three phases, A, B, and C, have been created with A at the top and C at the bottom. A change in the shape, size, mode, etc., of B would clearly require recalculation of B's particles, but would also change the fraction of the image covered by C, and so C's particles would need to be recalculated as well, whereas A is unaffected and thus does not require recalculation.

The blob shape is substantially more complex than the other three shape types because it is stored as a cubic Bezier spline, whereas the others are stored as graphics primitives (ellipse, rectangle) or polygons. Although the additional complexity is generally opaque to the user, in the overlap calculations it may become apparent. This stems from the blob shape being created identically to a polygon but then being smoothed, using the polygon vertices as anchor points (points through which the curve must pass). The smoothing algorithm creates two additional control points between each pair of anchor points (Shemanarev, 2009) to produce a smooth curve that passes through each polygon anchor point. Detecting overlap between blobs based on the Bezier curves would be computationally costly, particularly for numerous small particles. The overlap algorithm, therefore, treats blobs identically to polygons for this purpose. Thus, nearby blobs in a phase set to disallow overlap may, in fact, have slight overlap where the blob outline lies outside the equivalent polygon.

MODEQUIZ

ModeQuiz is a companion program to ModeMaker, but designed for students to use on their own to improve their mode-estimation skills. ModeQuiz is based on the same foundation as ModeMaker, but rather than make images to the user's specification, the program produces randomized images, and the user is asked to make estimates for each

TABLE I: ModeQuiz image-creation properties specified by random selection. If mean and standard deviation values are absent, then selection is random over the uniform distribution bounded by the listed extreme values. If mean and standard deviation values are present, then selection is random from a normal distribution with the given mean and standard deviation values and, in some cases, with additional constraints dictated by the given extreme values. In most cases, the extreme is a truncation: if the random selection from the normal distribution exceeds the truncation value, then the property adopts the truncation value itself. However, "abs" indicates that rather than truncation at the minimum, the absolute value of the random selection is adopted, and a fixed amount added, as indicated by the accompanying number.

Property	Mean	Standard Deviation	Minimum	Maximum
Number of phases	1.5	0.5	1	5
Mode per phase (if overlap is permitted)			2	90
Mode per phase (if overlap is prohibited)			2	40
Size			8	108
Standard deviation of particle sizes	Size/10	Size/2	0 (abs)	
Complexity			3	10
Standard deviation of complexity	0	2	0 (abs)	
Aspect ratio	2	3	1 (abs)	
Standard deviation of aspect ratio			0	aspect ratio × 2

phase. Once estimates are made, the correct answers are revealed, a score is assigned, and a new image is created. By noting the score, the student can gauge his or her improvement.

Randomization Settings for Images

Because ModeQuiz is based on the same foundation as ModeMaker, all the image specifications described above are present in ModeQuiz, but rather than being specified by the user, they are selected randomly. For some of the properties, the selection is performed randomly over a uniform distribution, but for others, the selection is made from a normal distribution characterized by a mean value and a standard deviation (Table I). Note that this mean and standard deviation are independent of the means and standard deviations specified in the ModeMaker user interface, and, in fact, in many cases, the value of these properties (e.g., the mean of the aspect ratio or the standard deviation of the particle size) is itself randomized using a normal distribution.

In addition to the properties listed in Table I, some additional constraints are imposed. The total mode may not exceed 98%; total modes very close to 100% require substantially longer calculation times because numerous particles are placed to cover the few uncovered portions of the image. The colors of each phase and the background are randomly chosen from a list of 13 predetermined, contrasting colors. The shape for each phase is randomly chosen from among the possible shapes. Particle overlap has a 50% chance of being allowed, but only if the phase's mode is less than 40%. Reentrants have a 50% chance of being allowed if the shape is a blob or polygon.

Scoring Algorithm

To allow the student to have a familiar and easy measure of his skill, a score is calculated in a manner similar to what might be expected on a quiz. Full credit (100) is given if the estimate for the phase is within 5 percentage points of the measured mode for the phase. The score decreases by 2 for each increased percentage point of distance from 5 away for the correct mode. The mean score for the session (i.e., the mean score for all estimates to that point) is updated after each set of answers are revealed. The mean score for the most recent five images is also displayed, so the student can judge his or her recent performance against his or her overall performance and thereby roughly estimate improvement over time (Fig. 3).

INSTRUCTOR/PROFESSIONAL ASSESSMENT

Because many instructors use their own estimates of modes as the "correct answers" in teaching mode assessment, it is illuminating to assess how proficient instructors and other professional geologists are at mode assessment. This assessment was performed over the Internet during October and November 2009. The modest number of participants, the limited control over the assessment environment, and the self-reporting of demographic data render the conclusions drawn from this assessment preliminary, but they are nevertheless enlightening.

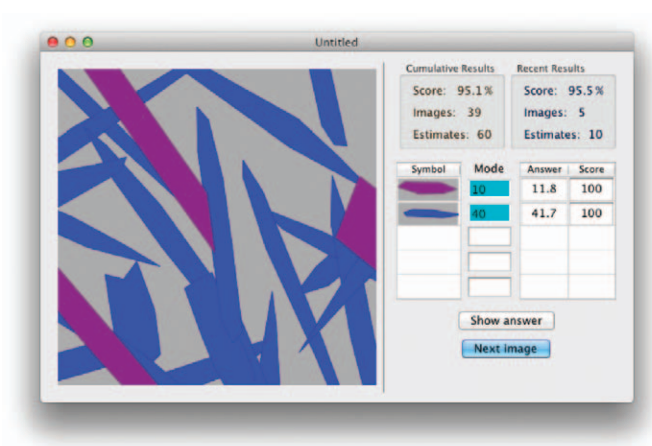


FIGURE 3: Sample ModeQuiz window, showing cumulative and most-recent scores.

Methods

A group of 59 geologists was recruited for this assessment via requests sent to two Internet mailing lists: the Teaching Petrology list (http://serc.carleton.edu/pipermail/petwksp/) and the Geo-Metamorphism list (http://www.jiscmail.ac.uk/lists/Geo-Metamorphism.html). This represents approximately a 6.5% response rate. Each respondent completed an online survey that contained an informed consent form, six questions designed to characterize how the respondent interacts with students and uses mode estimation, and 12 mode estimation tasks, each involving one of four components in one of three granitoid rock images (Fig. 4). The responses were kept anonymous by the survey system, the free, open-source "LimeSurvey" (http://limesurvey.org/) running on the author's own webserver.

Correct answers for the mode questions were determined using Adobe Photoshop. This image analysis method is based on an attempt to select all pixels in the image belonging to each phase in turn, then counting those selected pixels. The primary tool used in the method allows the interactive selection of colors. The user adds colors to the currently selected set until all pixels of the phase are selected, and no pixels of other phases are selected. In some cases, there is an overlap of colors (the same color belongs to different phases in different locations); this problem is more pronounced when the rock is not stained for plagioclase and K-feldspar. In such cases, the method errs on the side of selecting fewer pixels, ensuring that no pixels are part of multiple phases.

To interpret the results, a score was calculated for each of the 12 estimation tasks, for each respondent. The scores are calculated to be normalized to the mode of the component. This accounts for the typical ability of most users to better distinguish small modes from each other (e.g., it is easier to distinguish 5% from 10% than 50% from 55%). The score is calculated by the following:

$$s=100\left(1-\frac{|c-e|}{c}\right), \quad (1)$$

where s is the score, c is the correct mode, and e is the mode estimate. Note that it could be argued that this score equation fails to account for the renewed ease of estimation achieved as the mode of the component approaches 100%, but none of the modes in the survey exceed 50%. The 12 estimation scores for each respondent were averaged together to derive a final score for each respondent.

RESULTS

The results showed that the mode estimation skills of the respondents were only fair, and they varied little with any of the measured demographic traits. The mean score for all respondents was 65.9 with a standard deviation of 7.5. That score corresponds to a misestimate of 17.1 percentage points, given a true mode of 50%, or 1.7 percentage points given a true mode of 5%. The scores ranged from 46 (by a respondent who believed himself or herself to be "proficient" at mode estimation by the definition in Fig. 4) to 78 (by an "advanced beginner"; Fig. 4). The standard deviation of the 12 estimation scores for each respondent provided a measure of how precise the participant's mode estimation skills were. Those values ranged from 17.4 (by the highest-scoring individual) to a high of 40.0 (by an individual with a mean score of 57).

There was little, if any, correlation with demographic measures. The only correlation found was a weak correlation between the score achieved and the years spent since first using mode estimation [Fig. 5(a)]. A weak inverse correlation was observed between the score achieved and the years spent teaching mode estimation [Fig. 5(c)]. Neither of those relationships was significant at the 95% confidence level. No significant relationship was found between mode-estimation proficiency and the perceived proficiency of the respondent [Fig. 5(b)] or the frequency of his or her use of mode estimations [Fig. 5(d)].

Discussion

These results suggest that, although there are some individuals who are proficient with mode estimation, most are not, including many of those who perceive themselves to be proficient. It is notable, however, that not one respondent perceived him or herself to be an expert at this skill. The results suggest that there is no clear benefit to years of teaching or experience with mode estimation. One possible explanation for this is that, like our students, we receive little feedback as to the correctness of our mode estimates, and when we do receive that feedback, it is far removed from the moment of estimation. We fail to learn for the same reasons our students fail to learn.

Knowing our lack of expertise with mode estimation, we should be wary of using our estimates as the correct answers for teaching or assessing students in mode estimation. On average, our estimates will be further from the correct answer than we would accept from our students. Although some of us have better skills than others in this area, it is of note that our own perceptions of our skills fail to correlate with our actual skills in this arena.

MODEQUIZ ASSESSMENT

The results of two small experiments can gauge the effectiveness of ModeQuiz on student learning. They show that 30 min of ModeQuiz training is insufficient to produce improvement in mode estimation on real rocks but that 2 h of training may provide noticeable benefits.

Methods—Experiment 1

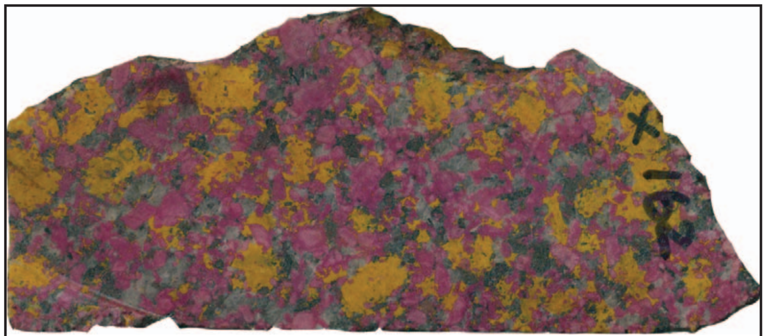
A group of 36 undergraduates was recruited from two nonmajor, sophomore-level, introductory geology courses. The students were randomly split into experimental and control groups based on whether the final digit of their student numbers was 0–3 (control) or 4–9 (experimental). Student numbers are assigned in order of registration, so this should be an effective source of randomness.

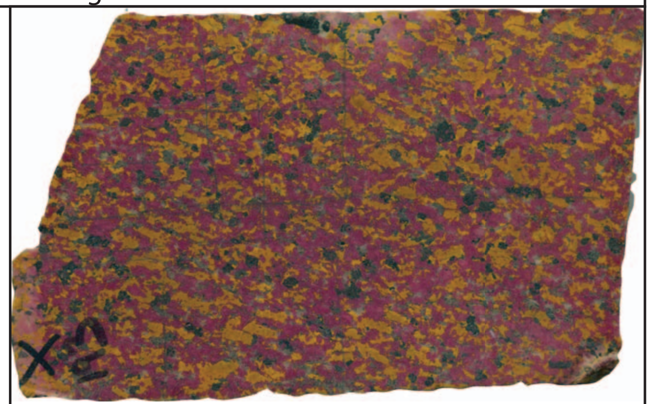
The primary experiment itself was very simple. Before the experiment, all students were asked to complete an assessment of their ability to determine modes. The assessment consisted of six granitoid specimens, five of which were stained for plagioclase and K-feldspar; the students were asked to estimate the modes of quartz, plagioclase, K-feldspar, and mafics + oxides (Fig. 6). They were given the reference sheet (Fig. 1) to use if they wished. To remove any mineral identification aspects of the skill, they were told the colors of the minerals (gray, pink, yellow,

1. What course or courses do you teach that involve mode estimation?
2. For about how many years have you taught the mode estimation skill?
3. In approximately what year did you begin using mode estimation skills yourself (presumably, but not necessarily, as a student)?
4. How often do you use mode estimation skills, either as part of your teaching, research, or professional work?
   - Never
   - A few times a year
   - A few times per month
   - A few times a week
   - Nearly every day.
5. How proficient are you at mode estimation?
   - Poor/Novice (within 20 percentage points of correct answer less than half the time)
   - Advanced Beginner (within 20 percentage points of correct answer more than half the time)
   - Competent (within 15 percentage points of correct answer more than half the time)
   - Proficient (within 10 percentage points of correct answer more than half the time)
   - Expert (within 5 percentage points of correct answer more than half the time)

6-7. Please estimate the modes of each of the following components in this rock image. Please do not use image processing tools to measure the modes.
   - Plagioclase (pink)
   - K-feldspar (yellow)
   - Quartz (grey)
   - Mafics/oxides (black)

8. Please estimate the modes of each of the following components in this rock image. Please do not use image processing tools to measure the modes.
   - Plagioclase (white)
   - K-feldspar (tan)
   - Quartz (grey)
   - Mafics/oxides (black)


Question 6 image


Question 7 image


Question 8 image

FIGURE 4: Questions asked to respondents in the "Instructor/Professional Assessment," intended to gauge the mode-estimation skills of professionals using mode estimation and those teaching mode-estimation skills. Note that respondents were provided with the mode-estimation reference sheet (Fig. 1). For the three mode-estimation questions, respondents were not asked to identify minerals but to estimate modes of colored components, with the colors of the feldspars dependent on whether the rock was stained or not.

and black, respectively, for the stained rocks, and gray, white, pink, and black for the unstained rock).

Next, all students were given mode-estimation training. The students in the experimental group were asked to train with ModeQuiz for 30 min with this guidance: "You should use the score for each estimate to gauge how you are doing and adjust your estimates accordingly. For example, if you find that you are regularly underestimating, try to adjust your estimates higher to compensate." The control group was designed to mimic the typical method of instruction in mode estimation: these students were instructed to examine four coarse-grained rocks and to make estimates and were then given the correct answers, along with the same guidance given to the experimental group. They were asked to spend as long as they could to calibrate their estimation skills, up to 30 min. Following the training, all students

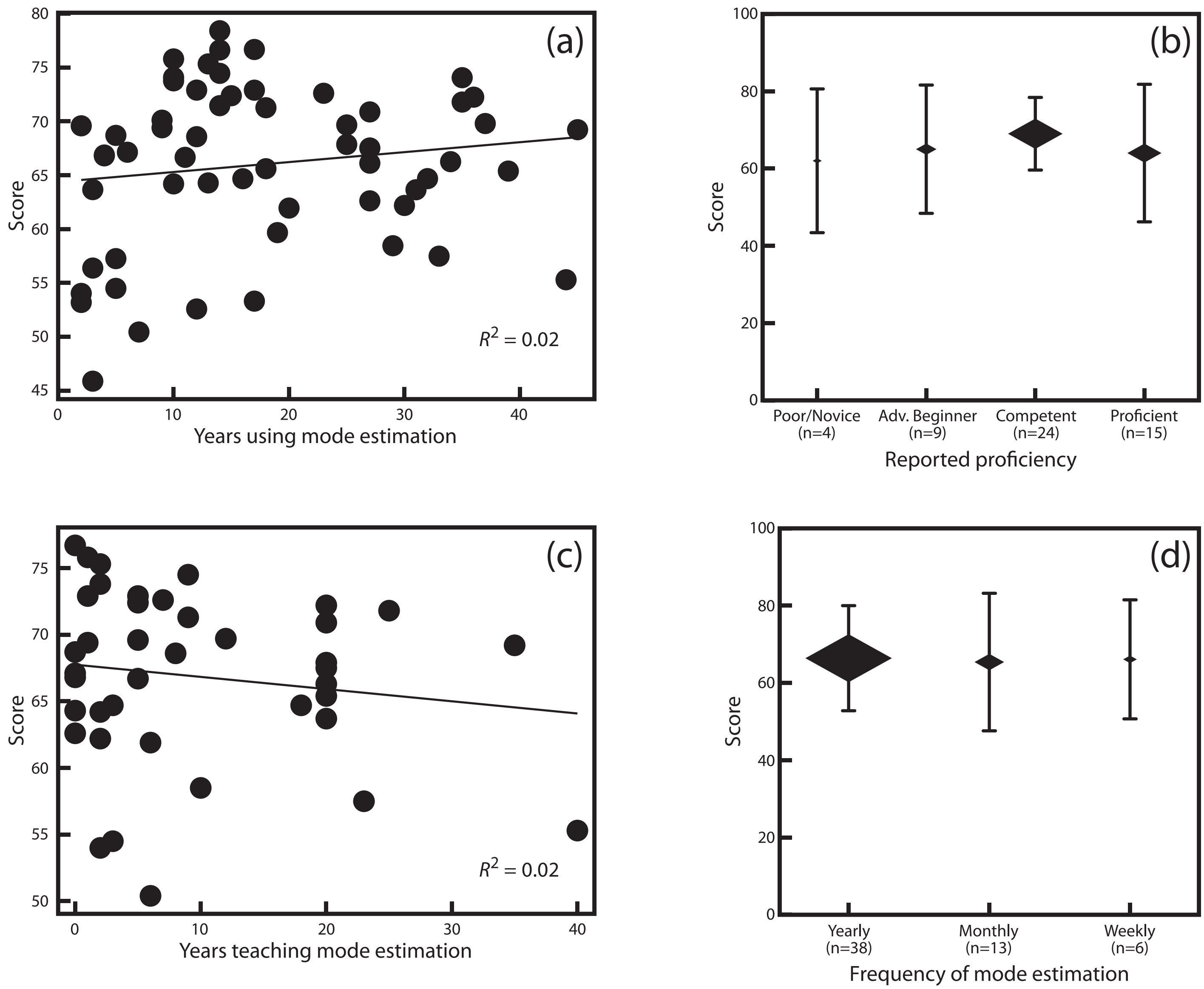


FIGURE 5: Summary of instructor/professional mode-estimation skills based on survey data. Score represents the accuracy of the respondents' estimations, averaged over 12 estimations on three rock images. A score of 100 was perfect. In (b) and (d), the symbol size is related to the number of respondents in the category (*n*), and the error bars are 2σ. In (a) and (c), the best-fit line through the data is shown.

retook the assessment, to gauge any improvement in their estimation skills.

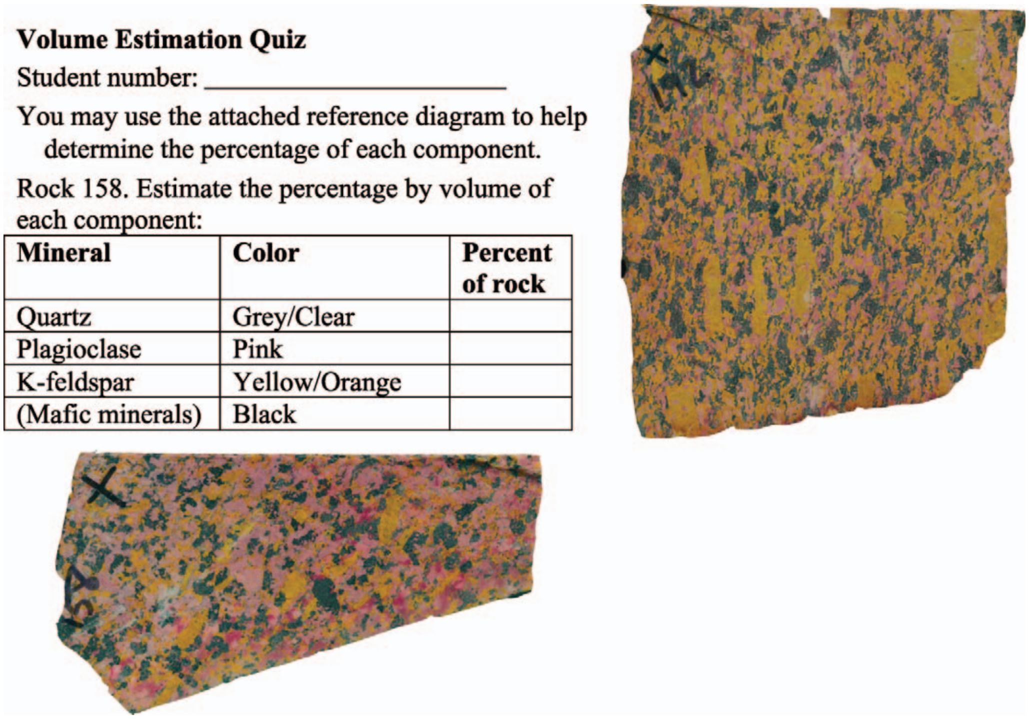
In an effort to enhance the effectiveness of the training and the assessment, prizes ($30 gift certificates) were offered for both the top scorer during the ModeQuiz training, and the top scorer in the posttraining assessment. All students received a small amount of extra credit in their class for participating.

Methods—Experiment 2

In a follow-up experiment, 18 students in a senior-level, advanced petrography course were studied. Mode-estimation skill is part of that course, and the students knew they would be evaluated on the skill eventually (but not as part of the experiment itself). The protocols were similar to the first experiment, with these exceptions: (1) students were asked to spend 2 h on either the rock training or the ModeQuiz software; (2) no prizes were offered; (3) the students had 2 wk to complete the training; and (4) students self-selected into the experimental or control groups, based primarily on whether they had home access to a Macintosh computer to run the program (suitable computers are available on campus, but few students chose to run the program on them for 2 h).

Results

The results of the first experiment weakly confirmed the hypothesis that 30 min of ModeQuiz training enhanced mode-estimation ability for real rocks beyond that provided by traditional training. For two specimens (188, 192) the values for "quartz" and "mafics" were combined because it was clear from the results that some students viewed the



Volume Estimation Quiz

Student number: ______________________

You may use the attached reference diagram to help determine the percentage of each component.

Rock 158. Estimate the percentage by volume of each component:

Mineral	Color	Percent of rock
Quartz	Grey/Clear	
Plagioclase	Pink	
K-feldspar	Yellow/Orange	
(Mafic minerals)	Black	

FIGURE 6: ModeQuiz assessment, with the question format shown (remaining five questions were similar). Students were asked to estimate modes in real rock specimens; two of those images are shown here. The remaining three were used in the instructor survey and are shown there (Fig. 4). All five specimens were 7–15 cm long.

dark grey matrix as quartz, estimating modes of 30%–40%, whereas other viewed it as mafics, producing quartz estimates of 0%–5%. In addition, the results for one student were discarded because the student's ModeQuiz score was a clear outlier (66%, compared with a mean for the other students of 92% ± 3%). The estimates made on the pretests and posttests were compared with the correct values (determined as described in "Instructor/Professional Assessment Methods" above) to derive an estimation error for each component of each rock before and after the training. The difference between the posttest error and the pretest error was the improvement. Dividing the improvement by the mode of the component gave the relative improvement. The relative improvement values were averaged across components to derive mean relative improvement values for each student (Table II). The experimental and control groups' results were averaged to derive mean relative improvement values of 0.01 ± 0.08 for the control group and 0.01 ± 0.07 for the experimental group. Statistically, the experimental group was indistinguishable from the control group, with an effect size (from Cohen's *d* statistic) of 0.02.

Other tests were performed to determine whether there were correlations between relative improvement and other factors. Male students performed better (0.04 ± 0.19) than did female students (0.004 ± 0.21), but these values were also indistinguishable within uncertainty. In a separate pilot study, a correlation between student grade point average and mode-estimation performance was revealed, but that correlation was supported only weakly in this data set ($R^2 = 0.02$).

The results from the second experiment were more significant statistically. The same data treatments were applied (combining quartz and mafics for samples 188 and 192). Students reported the amount of time they spent training, and because the experimental conditions were less constrained, not all students trained for the same amount of time. Those that did any training at all were separated into a control group (who trained with rocks, $n = 5$), a "good" experimental group (who trained for at least 2 h with ModeQuiz, $n = 6$), and a "poor" experimental group (who trained with ModeQuiz for less than 2 h, $n = 4$). The good experimental group had a relative improvement score of 0.02 ± 0.05; the poor experimental group has a relative improvement score of −0.01 ± 0.02, and the control group had a relative improvement score of −0.01 ± 0.07 (Table II). These values correspond to effect sizes (from Cohen's *d* statistic) of 0.75 for the difference between the control groups and the good experimental group and 0.72 for the difference between the good and poor experimental groups.

Discussion

The results from the experiments show that a small amount of focused effort using ModeQuiz did not improve students' mode-estimation skill for real rocks, but more-substantial amounts of training were correlated with significant improvements in mode-estimation skill. The effect size for the results of the first experiment, 0.02, was a negligibly small effect (Cohen, 1988). The effect sizes for the second experiment, 0.75 and 0.72 were large effects (Cohen, 1988) and suggest that 2 h of training with ModeQuiz offered real benefits over either training with rocks (from 0.2–1.0 h) or training with ModeQuiz for less time (from 1.0–1.5 h). Note that one advantage of ModeQuiz

TABLE II: Relative improvement values for each student. Categories in Experiment 2 are described in text. Each value is calculated from an average of pretraining and posttraining assessments on 19 components in five specimens.

	Experiment 1		Experiment 2		
	Control	Experiment	Control	"Poor" Experiment	"Good" Experiment
	−0.03	0.08	−0.07	−0.01	0.08
	0.05	−0.04	−0.01	0.02	0.04
	0.10	0.07	0.08	−0.02	0.00
	0.05	−0.04	0.04	−0.01	−0.06
	−0.06	0.03	−0.07		0.06
	0.07	−0.05			0.01
	0.12	0.07			
	−0.06	−0.06			
	0.02	0.13			
	0.05	−0.04			
	−0.24	−0.09			
	0.05	0.06			
	−0.01	0.00			
	0.01	−0.11			
	−0.03	0.09			
	0.09	0.01			
	−0.02				
	0.04				
Mean	0.01	0.01	−0.01	−0.01	0.02
Std. Dev.	0.08	0.07	0.07	0.02	0.05

is that students can profitably train with it for far longer than they can with a modestly sized set of real rocks.

Notably, in the first experiment, most students showed improvement within the program after 30 min of ModeQuiz use: the program tracks the cumulative score as well as the recent score, and by the completion of the training session in the second phase of the experiment, the recent score was typically higher than the cumulative score, suggesting improvement (these recent-score values were not systematically recorded, so the data are not available). This result also suggests that students were indeed learning the skill in the experimental environment. The disconnect between improvement in mode estimation performance within ModeQuiz and lack of improvement in the rock-based assessment in the first experiment may be related to the nonrealistic nature of the program's images.

The program shows the user images that are not real rocks and, in many cases, do not resemble rocks in hand specimen, although they bear a stronger similarity to thin-section views. It may, therefore, be the case that assessments with thin-sections would show a stronger effect than hand-specimen assessments. Nevertheless, 2 h of training with these images appears to have substantial benefits for the student, even using hand-specimen assessments.

There are a number of caveats to the conclusions drawn from the second experiment. Because the students self-selected into the experimental and control groups, there is ample opportunity for bias. Some students worked less at training than others did, which might be related to skill or motivation, again, providing an opportunity for bias. No student in the control group trained longer than an hour, whereas the good experimental group trained at least 2 h; that training time factor cannot be separated from the training method factor in the results, so it may be that 2 h of training time produces benefits, regardless of training method. Finally, the number of students in the experiment was very small, and a larger experimental population might have revealed a different result.

CONCLUSIONS

A difficulty associated with training students in the skills of mode estimation is that few practicing geologists are well trained in mode estimation; this is borne out by the survey results (Fig. 5). This weakness prevents us from using our own estimation skills to effectively teach students or to assess student abilities in this area. Recognizing this weakness will allow us to compensate for our own lack of training, to effectively train our students in mode estimation, and to assess the state of their knowledge.

We must develop effective training methods. By using ModeMaker, we can develop sample images with known modes on which students can be trained. Once they have mastered artificial images, we should train them on real rock specimens, but knowing our weakness in mode estimation, we should resist using our estimates as the correct answers, and instead, to effectively train students, we should use image-processing tools or point counting to determine the correct answers. ModeQuiz may be a useful adjunct to this training.

We must fairly assess student estimation efforts. We can use ModeMaker to create sample images with known modes on which to test students. If we use real rocks, then, as above, we must recognize our limitations and use image-processing software or point counting to derive the correct answers, rather than using our own estimates as the correct answer. If we can do a more effective job training our students, then when they take teaching positions, they will be better able to use their estimation skills for teaching purposes.

FUTURE DIRECTIONS

Development is ongoing for both applications presented here, with two features in particular demand: better images and more accessibility. Firstly, future versions of the program will create better approximations for real rocks by substituting for the current strategy of shape placement, a simulated nucleation and growth process, with ideal shapes for each mineral explicitly coded, and with growth order and timing specified (by the user for ModeMaker; semirandomly for ModeQuiz). Secondly, versions of the program that run on tablets and smart phones will be created to spur wider adoption.

OBTAINING MODEMAKER AND MODEQUIZ

These programs were written in Objective-C, using the Cocoa frameworks on Mac OSX. They do not run on Windows or Linux, except perhaps in emulation. Current versions require Mac OSX version 10.5 or higher running on an Intel central processing unit (CPU) and have been tested on version of Mac OSX through 10.7.3. Development and improvement is ongoing, but as of this writing, ModeMaker has reached version 2.1 and ModeQuiz has reached version 1.61. These versions of the programs and source code are freely available for use and modification under the GNU General Public License v3.0 (http://www.gnu.org/licenses/gpl.html). They are posted online as supplementary files associated with this article. (Supplementary files are available at dx.doi.org/10.5408/11-254S1.) The most current versions will be available indefinitely at http://davehirsch.com, as are older versions that run on Mac OSX 10.4 and on a PowerPC CPU.

Acknowledgments

Thanks to all those who tested early versions of the software and offered constructive feedback. Thanks especially to Western Washington University students who participated in the experiments. Thanks as well to Apple Computer for providing such a smooth and elegant language and API (application programming interface) with which to easily create powerful software (ModeMaker version 1.0 was completed in 6 d). Helpful comments from an anonymous reviewer improved the manuscript substantially.

REFERENCES

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. London, UK: Routledge Academic. 567 p.

Delesse, A. 1866. Procédé méchanique pour déterminer la composition des roches. Paris, France: F. Savy. 8 p.

Dilks, A., and Graham, S. 1985. Quantitative mineralogical characterization of sandstones by back-scattered electron image analysis. *Journal of Sedimentary Petrology,* 55:347–355.

Galehouse, J. 1971. Point counting. *In* Carver, R.E., ed., Procedures in sedimentary petrology. New York: Wiley-Interscience, p. 385–407.

Kulik, J., and Kulik, C. 1988. Timing of feedback and verbal learning. *Review of Educational Research,* 58:79–97.

Le Maitre, R., Streckeisen, A., Zanettin, B., Le Bas, M., Bonin, B., and Bateman, P. 2005. Igneous rocks: a classification and glossary of terms. Cambridge, UK: Cambridge University Press. 256 p.

Marschallinger, R. 1997. Automatic mineral classification in the macroscopic scale. *Computers and Geosciences,* 23:119–126.

Philpotts, A.R. 1989. Petrography of igneous and metamorphic rocks. Prospect Heights, IL: Waveland Press. 192 p.

Shemanarev, M. 2009. Interpolation with Bezier curves. Available at http://www.antigrain.com/research/bezier_interpolation (accessed 4 April 2012).

Tinkham, D., and Ghent, E. 2005. XRMapAnal: a program for analysis of quantitative X-ray maps. *American Mineralogist,* 90:737–744.