


# COFFEE GRADING ANALYTICS

## Testing the Tester

by K.C. O'Keefe

photos by K.C. O'Keefe

illustrations by Jonathan Landa



ONE SPRING MORNING in 2008, I looked at the cupping lineup in front of me and felt slightly uncomfortable. Sitting before me were nine origin samples for potential purchases: five conventional Sumatras and four Nicaraguan direct-trade pre-ship samples. My three cupping comrades and I faced the same task: Evaluate all of these coffees to the roasting company's 85-plus-point precision. We had time for only one shot at the cupping table. Could we do it? I tried to shake off my nerves and jumped in like a good soldier storming the beach.

Two years later, on the other side of the planet, 38 Licensed Q Graders

in four labs gathered at the start of the harvest season. This session was the first of many cupping sessions in which they would evaluate a nation's worth of coffee for the Ethiopia Commodities Exchange (ECX). Again, the same nagging feeling arose. Workneh Mulat, the ECX's quality control manager, urgently posed the questions: How do we ensure that each lab protocol and each grader is consistently calibrated with the others? How can we assure the validity of our cupping results?

Every day, our industry makes thousands of decisions based on a Supreme Court of cuppers who are asked to impartially and accurately judge the labor results of millions

of people. We constantly encounter miscalibration of point and relative quality values of coffees, which frustrates both specialty coffee sellers and buyers. So how do we ensure ongoing accuracy, repeatability and calibration between cuppers? How can we be self-assured that our 84 really is not an 87? And if grading is relative, is it relative between all cuppers? Or worse, is it relative for the same cupper from one day to the next?

How can our industry assure the producing world that the Supreme Court of cuppers is just? Does the cupper in your lab produce valid results?

*continued on page 24*

**PACKAGING...**



**...SIMPLIFIED.**

Pre-Made Bags • Stand-Up Pouches •  
Patented One-Way Valve • Quad Seal Bags •  
Side Gusset Bags • Rollstock • Jute Bags •  
Specialty Tins • Equipment •  
Reusable Packaging • Totes



T. 425.455.1128 || F. 425.990.8582  
[www.pacificbag.com](http://www.pacificbag.com) PACKAGING...SIMPLIFIED.



**STEAMING PITCHER RINSER** BY ESPRESSO PARTS

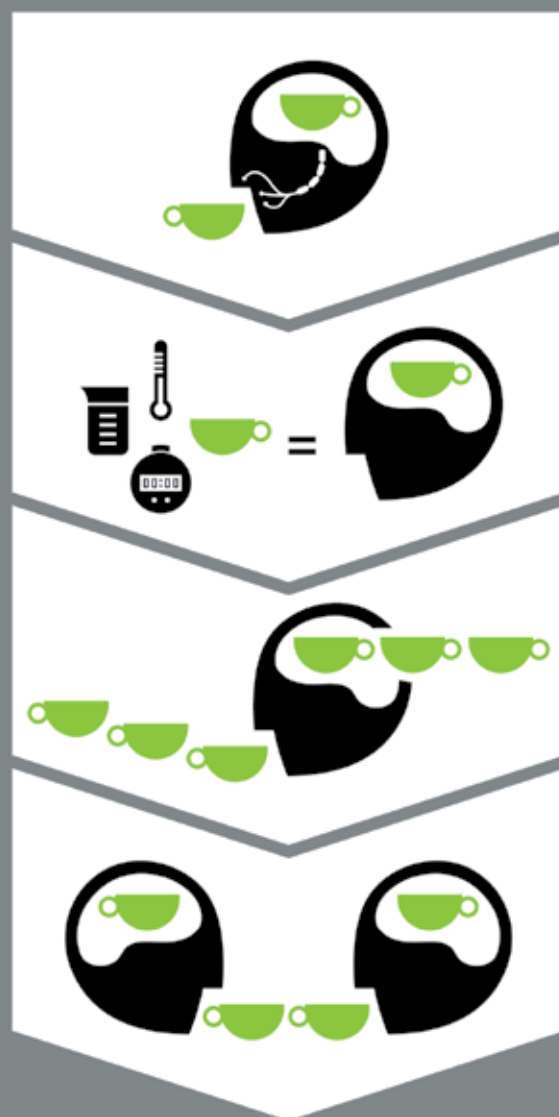
Rinsing steaming pitchers between each use is not only sanitary preparation practice, it is also beneficial to the proper production of finely textured milk beverages.

The Espresso Parts Steaming Pitcher Rinser is the answer to effectively rinsing your pitchers and conserving water. Our steaming pitcher rinser can be configured into virtually any application. Available in 6" x 6" x 4", 6" x 10" x 4", and 24" x 7" x 7/8" with built in water and drain connections. Adaptable to most any espresso bar. Comes with easy to follow installation instructions.

**ESPRESSO PARTS** { [ESPRESSOPARTS.COM](http://ESPRESSOPARTS.COM) }  
1(800)459-5594 - (360)357-7781  
600 4th AVE EAST OLYMPIA WA 98506

## COFFEE GRADING ANALYTICS

Welcome to coffee grading analytics, the study of the process and results of the systems and individuals that judge coffee quality. Grading analytics can be divided into two categories: system analytics and grader analytics. System analytics is the study of the validity of a grading system and methodology, for example, the Specialty Coffee Association of America (SCAA), Cup of Excellence (COE) or an independent system. Grader analytics is the study of the validity of a coffee tester's results when using a pre-established protocol. Grader analytics focuses on three core metrics: ability, repeatability and calibration. This article will focus on analyzing ourselves as individual quality testers of single-origin green coffee purchases.



## 1 ABILITY



Q Grader testing makes several evaluations of a cupper's ability to taste and smell.

The exams start with the familiar sensorial skills test that measures sensitivity to differing levels of acidity, sweetness and salt. Next, the exams move on to olfactory tests using Le Nez du Café aroma vials. Finally, the cuppers arrive at coffee-based tests by completing five triangulations and one organic acids matching-pairs exam. In reality, these tests not only evaluate the sensorial ability of the individual, but the individual's ability to concentrate and remember the sensations long enough to record them in written form (a basic requirement for professional cuppers). Most of the objective ability tests are fairly simple to replicate in any private lab, or even your personal kitchen, and do not depend on the quality of the coffees used.

## OBJECTIVE ABILITY TESTS

- 1 Sensory Skills Test
- 2 Olfactory exam (Le Nez du Café)
- 3 Triangulations
- 4 Match Pairs

So what's the best test for evaluating a specialty coffee taster? According to CQI's executive director, Ted Lingle, "Triangulations are the best test in evaluating whether or not a specialty coffee cupper can differentiate one coffee from another." Q Graders must pass five triangulations with 83 percent or higher accuracy. Basic objective ability is the first question we should face in evaluating our cupping abilities.

*continued on page 26*

**THE ELEGANT WAY TO MOVE COFFEE >>>**



**'Cableflow'**  
Tubular Cable Drag Conveyors

- Proven over 40+ years
- Cable, chain and solid link options
- Cable models offer long cable life
- 24/7/365 reliability
- Multiple inlets and outlets
- No filters or cyclones

Spiroflow, serving the Coffee industry for over 25 years!  
More Efficient, Labor Saving, Safer, and Contamination Free! Call us today.

**SPIROFLOW** GLOBAL POWDER HANDLING SOLUTIONS  
TEL: 866-876-7931  
E-MAIL: [info@spiroflowsystems.com](mailto:info@spiroflowsystems.com)  
WEB: [www.spiroflowsystems.com](http://www.spiroflowsystems.com)

as a matter of fact  
**LOOKS ARE EVERYTHING**



**pumpSkins**  
RESURFACING SYSTEMS

**Make a good first impression with pumpSkins!**  
We're branding experts offering all POP materials including equipment overlays, rack signage, flavor labels, window decals, signs and more.

**1-877-994-4600 • [www.pumpskins.com](http://www.pumpskins.com)**

## 2 REPEATABILITY

A thermometer's repeatability is easy to test. Simply put the probe in a cup of room-temperature water and read the results. Repeat the process, and the temperature should read exactly the same. If it does not, either the water temperature has changed or the thermometer should be destined for the Dumpster. One can only be confident the water temperature didn't change if the water temperature is controlled. Variables such as a delay in the time of reading, sunlight on the glass or having the cup on a hot plate would change the actual water temperature.

Discipline in controlling variables is the first step in producing repeatable results. Lack of controlling the variables is the top reason for invalid cupping results. I remember being corrected by George Howell several years ago when he measured the water temperature in the lab before cupping. "If it's boiled, that's enough, right?" I asked Howell, the owner of Terroir Coffee Company and a co-founder of Cup of Excellence.

"Absolutely not," he replied. "Measure each of those boilers and see what temperature they give." Howell was right; although they had all come to a boil within 60 seconds of one another, the three boilers had a deviation of 7 degrees F (4 degrees C). Worse yet, many labs are not equipped with thermometers for measuring



water temperature, making it impossible to follow industry water standards as recognized by the SCAA.

The concept may seem simple enough, but to be repeatable, we must control the controllable variables. The roast, water and measurements are all easily controlled and simply demand discipline in the lab. Unfortunately, much of our industry fails to equip itself with the precision instruments necessary to create repeatability. For example, most sample roasters do not have gas pressure controls (precision valve and pressure readout), yet coffees can easily demonstrate a two-point or more cup variation solely due to different roast profiles roasted to the same color reading.

Once cuppers who work in equipped labs are consistently disciplined in their protocols, they can measure their ability to produce repeatable results. The simplest way to measure repeatability is to blindly evaluate a single roast sample twice in the same day and compare the results. The closer the points (and/or

words) are to each other, the lower the point deviation and the higher the repeatability.

Let's stop for a moment to discuss the concept of graduated sensorial accuracy. Acceptable sensorial precision (exact cupping scores) have graduated relevance in coffee's sensorial analysis. In my experiences with commercial-scoring coffees (70-80 points), precision is not required to be "accurate" or "calibrated"; the range of calibration may be as high as +/- 4 points. The lower the grade of coffee, the greater the range of acceptable sensorial accuracy. On the inverse, sensorial grading precision is more important and demands greater precision the higher the grade (cupping score) to be considered accurate. Specialty coffee cuppers are expected to have the greatest level of "accuracy," yet in reality many cuppers in the commercial business have greater training, tighter calibrations and tighter standards than do many specialty roasters who lack formal sensory-science training.

During cupping courses last summer, I had the opportunity to put the repeatability evaluation methodology to the test by appraising three groups of Peruvian specialty cuppers who were preparing

*continued on page 28*

## CONTROLLABLE VARIABLES

- 1 Roast time vs. temperature curve
- 2 Roast color
- 3 Roast-to-grind elapsed time
- 4 Coffee-to-water ratio
- 5 Grind particle size
- 6 Grind-to-fragrance elapsed time
- 7 Blind codes vs. disclosed origins
- 8 Water quality (TDS, pH and hardness)
- 9 Water temperature
- 10 Pour-to-crust-breaking elapsed time
- 11 Quantity and form of stirs
- 12 Number of passes vs. elapsed time

The Original **Green** Coffee Sleeve.

**JAVA JACKET.**  
Supporting a greener future for our planet.

1-800-208-4128 503-281-6240 Fax: 503-281-6462  
info@JavaJacket.com www.JavaJacket.com

*Is Yours Made in the USA?*

*Ours Is.*

Astra manufactures 18 different models of espresso machines using only the finest components. A full line of unique, dependable, state-of-the-art espresso machines at an exceptional value.

2 YEAR WARRANTY ON PARTS  
EXCEPT HIGH WEAR-AND-TEAR ITEMS

**Astra**  
MADE IN THE U.S.A.  
www.astramfr.com

877/340-1800 818/340-1800



for their Q exams. I thought it was reasonable to expect that a trained specialty coffee cupper using the SCAA format should be able to repeat a sample's total point score within a 1.5-point deviation. Thus, I established a pass/fail protocol of +/- 1.5-point deviation to be considered repeatable with the same roasted coffee cupped twice on the same day. To test the cuppers, I placed six samples on the cupping table in the morning, and then repeated the exercise in the afternoon, setting the coffees in distinct positions. In each class, the cuppers were tested on five separate days, cupping coffees from various origins.

Of the 24 cuppers who surpassed Q evaluation standards of more than an 80-percent average on the objective abilities tests (sensory skills, olfactory and triangulation), only four of them (17 percent) were able to repeat their cup scores within the +/- 1.5-point deviation more than 70 percent of the time. Eight out of the 24, or 33 percent, couldn't repeat their results 50 percent of the time. These results seem to shed some doubt on this particular group of cuppers. One could even deduce that the majority of these cuppers are unreliable. Maybe our expectations were too high?

I chose the +/-1.5 deviation to reflect ranges that specialty buyers expect for coffee deliveries. It's common for specialty roasters to contract 85-plus- vs. 82-plus-point

coffees. Perhaps an 84 could pass as an 85, but an 83 would be rejected for the 85-plus contract. Thus my opinion remains that specialty coffee cuppers could be considered calibrated with themselves if they can reproduce cupping scores within a +/- 1.5-point range. Whatever one determines is the acceptable deviation range, the methodology of repeatability evaluation is easy to execute. And if we cannot produce repeatable results, then it's unjustifiable for us to require producers to deliver coffees within a precise point range.

Sensorial analysis specialist Paul Songer has developed many techniques for testing cuppers. He explained the test used by the Cup of Excellence. "The best test for evaluating cuppers is actually having them cup the types of samples they will be expected to cup, under the conditions they would be expected to cup, and include

*continued on page 30*

## +/- 1.5 REPEATABILITY QUOTIENT EVALUATION

- 1 Take six distinct origin samples
- 2 A second person sets up the cupping table and conceals the identity of each sample
- 3 Cup the table, filling out your usual cupping form
- 4 Repeat steps 2 and 3, while changing the positions of the samples
- 5 Reveal the identity of each sample
- 6 Compare your results
  - Each result within +/- 1.5 receives a point
  - Divide the number correct by the total number of samples
  - 6/6 within +/- 1.5 = 100 percent repeatability; 5/6 = 83 percent repeatability
- 7 For greater evaluation accuracy, this test can be repeated several times

**ABSTRACT ART**

**OUR MASTERPIECE** Mountain Water Decal

TOLL FREE: (888) 416-2030 E: DESCAMEI@PRODIGY.NET.MX W: WWW.DSCAMEX.COM

**NUOVA RICA MBI USA** World's #1 wholesaler of espresso machine & grinder parts.

In addition to our large inventory of parts, we also have a complete line of espresso grinders, hoses, fittings, cleaners and barista tools.

*The One Stop Shop*

**Do you dream in COLOR?** Barista accessories available in your favorite color...

Tampers • Cleaning Brushes • Scoops  
Knockboxes • Cocoa Shakers

[www.nuovaricambi.us](http://www.nuovaricambi.us)  
(916) 315-2535 • [sales@nuovaricambi.us](mailto:sales@nuovaricambi.us)

**ai allen flavors inc.**

**Makes The World Taste Better**

- Hundreds of natural gourmet coffee roasting flavors
- Beverage Development
- Natural Coffee and Tea Extracts

Ph.: 908-561-5995 • Fax: 908-561-4164  
23 Progress Street • Edison, NJ 08820  
[www.allenflavors.com](http://www.allenflavors.com)

replications,” says Songer, who’s the COE’s quality control lead and technical advisor. “For the COE national juror test, we have [cuppers] do three sessions of eight samples each, which they then score using the COE form. We carefully select a range of coffees, and there are three of the same samples that appear on each table in different positions for the replications.”

Songer says the No. 1 cupping error is the lack of replication of results. “A close second error,” he adds, “is lack of coding of samples. No matter how good a cupper one is, there will still be a psychological suggestion if you know the coffee is from such-and-such an origin, region, farm or even importer.”

So what is your repeatability quotient? If you blind-cup six coffees in the morning and the same six in the afternoon, are your scores acceptably similar? Are they within one point of each, or five points? Can you evaluate the same coffee twice and come up with final results within +/- 1.5 points?

## 2 CALIBRATION

Once we have established that a cupper has the basic ability and discipline that produces repeatable cupping results, we can finally

move on to the most complex question of cupper calibration.

A calibrated cupper can be defined as one who has the repeated ability to match the results (scores and/or words) of an established group or individual. As with repeatability, we must ask the question, how close

is close enough? In many specialty coffee trades, it is expected that cupping results should be within one or two points’ maximum deviation of the established contract. If we make buying decisions based on a razor-sharp point (85+), then our provider partners should expect us to have a repeatability quotient equal to, if not better than, that precision. Without establishing our repeatability, our calibration with one another is out of the question.

The next question we ask in calibration is, calibrated to whom? The answer inevitably defaults to calibration with the buyer(s). Thus specialty coffee roasters’ cuppers are the Supreme Court of the land, with whom all sellers should calibrate.

For the past several years, I have been analyzing our Peruvian Transparency Contract buyers’ “cupper calibration” against that of the producers/exporters. To my surprise, in some cases we find more than 40 percent miscalibration. What is more troubling is that this miscalibration has not only been with producers, but also between the buyers themselves. Producers often ask, “Why are our results so different from our buyers’ results? We are Q

## CALIBRATION OPPORTUNITIES

- 1 Practice cupping daily/weekly
- 2 Sample exchange
- 3 Cupper association
- 4 Cup of Excellence
- 5 Coffees of the Year

Graders; we use defined protocols; we use bottled water.” Significant miscalibration is certainly a threat to any direct-trade relationship and, whether valid or not, remains the top criticism of specialty coffee buyers.

Limiting the communication stream to numbers alone stifles calibration building. The more we can communicate in valuable descriptive language the “why” behind our numbers, the more we will gravitate toward the same opinions and points regarding specific coffees. Without dialogue, consensus is very difficult to obtain. This is certainly understood by many in the direct-trade movement who meet face to face with the cuppers in faraway places and follow up on those travels with the exchange of cupping reports.

The most common reason for miscalibration is a lack of continual practice. If you have not been cupping for a while, you are sure to notice that your ability to cup declines within a two-week window. Cupping is much like playing a musical instrument; in two weeks’ time I haven’t forgotten the notes, but my fingers are slow in getting into position. Professional cuppers should try to perform sensory evaluations at least three times a week to maintain their abilities. Songer affirms, “I would rather have on my panel a less experienced person who has been regularly cupping several times a week for the previous several weeks, than a more experienced person who hasn’t picked up a spoon in a month.”

Another common reason for lack of calibration between serious cuppers is the effect of isolation decalibration, where one always cups with the same group of cuppers and/or cups a restricted selection of similar coffees types (from a limited geographic production area or processing type). This is a problematic reality for most cuppers. At origin, cuppers are usually limited to coffees from

*continued on page 32*



*Great People Working Together...  
Bringing You Great Service and Great Coffee!*



*When the Quality of Life and the  
Quality of Coffee are Both Important*

**Coffee Holding Company**

3475 Victory Boulevard  
Staten Island, NY 10314

**Tel: 800-458-2233**

[www.coffeeholding.com](http://www.coffeeholding.com)

**Organic Products Trading Co.**

*A Division of Coffee Holding Company*

P.O. Box 2994  
Vancouver, WA 98668

**Tel: 888-881-4433**

[www.optco.com](http://www.optco.com)

[www.cafefemenino.com](http://www.cafefemenino.com)

an established geographic area or processing type. At destination, cuppers often work by themselves, or at best with the same one or two cuppers in their business.

A sample exchange is the simplest way to combat isolation decalibration. Find a respected cupper and trade green coffee samples. Independently roast and cup them in your labs using the same protocols and compare notes. This can be done each year with your origin partners prior to the harvest, without the pressure of the buying/selling relationship. In Peru, for example, cuppers have been invited to join a cupping association through which samples will be exchanged and members will cup together at least twice a year.

The best way to calibrate with others is by cupping together. This highlights the tremendous effect that the Cup of Excellence and other competitions have had on cupper calibration. Perhaps COE's greatest effect has been bringing 20 or more cuppers together for a three- or four-day period. Rightfully the COE's focus is on finding the best coffees, but indirectly any competition is a cupper's calibration workshop.

The Coffees of the Year (COTY) program at the SCAA conference and cupping exercises at the Roasters Guild Retreat are additional opportunities to calibrate with peers. Team up with a respected cupper and compare notes from a side-by-side cupping. The key of any calibration exercise is to get past the buyer/seller or expert/beginner relationship and honestly cup coffees.

Of course, the first step toward becoming a calibrated cupper is to accept our own fallibility and realize that we have a need to learn from each other. Once that is acknowledged, it follows that cuppers will be more circumspect in their analysis, more willing to reconsider, more willing to give a coffee a second trial if there is even a whisper of a doubt. Independent overconfidence is one of the most disabling handicaps a professional cupper can possess.

As I ponder our quest for making roasting and buying decisions based on cupper point precision, I am reminded of the Cup of Excellence. A COE competition is a months-long process of hundreds of samples and panels of national and international cuppers who roast and cup each sample at least five times. A single COE sample has been roasted five times, cupped in two national rounds by at least 10 cuppers, and cupped in at least three international rounds by at least 15 cuppers—providing rock-solid support for a consolidated, precise cupping score.

Have we in effect taken the COE model and tried to apply it to our individual companies? We are intellectually dishonest and justifiably unfair if we believe we can replicate that level of precision in our own private labs with a sample roaster without gas controls, undisciplined protocols and unqualified graders. Obviously we do not have the independent ability to invest in a COE for our company alone, but a gas manometer on our sample roaster and a repeatable cupper are surely within our reach.

Measurable results are critical to be able to consider quality control valid. Unless we evaluate ourselves as cuppers and prove we are valid quality control instruments, we have no right to ask producers to deliver coffee to sensorial cup precision or ask our roaster to change the roast profile. As with any quality control instrument, the cupper must be verified as having sufficient sensorial ability. Once sensorial ability is confirmed, protocols and process discipline can be audited. It is impossible to consider a cupper repeatable or calibrated prior to proving he or she has sensorial ability and basic lab discipline.

Scattered across the globe are thousands of individuals dedicated to sample roasting, weighing, grinding, slurping, spitting and evaluating specialty coffee. In many cases, we are in our own hamster cages, running the fixed cupping wheel. We report back to each other our results, which place thousands of containers on the water and distribute billions of dollars to rural coffee hills. Unless we mature in the validity of our cupping results, I'm afraid the producers of the world will assume we are all just running in place.

K.C. O'KEEFE has been a cupping instructor since 2003 and is currently a Q instructor for the Coffee Quality Institute. He is the founder of Café Verde, a Peruvian specialty coffee company located in Lima, Peru. He can be contacted at [kc@cafeverdeperu.com](mailto:kc@cafeverdeperu.com).

