

# Controversial Wine Judging Study: The Real Story



© AFP | Bottles laid out for tasting at the Concours des Grands Vins de France

Retired statistics professor Robert Hodgson has owned Fieldbrook Winery in northern California for 35 years. Several years ago he released a study which showed that wine judges at the California State Fair gave dramatically different scores to the same wine when tasting it blind on two different occasions. The study has popped back into the mainstream media this year, cited as proof that wine tasting is "junk science" or "bullshit." **W. Blake Gray** asked Hodgson what the study really means.

Posted Wednesday, 17-Jul-2013

## **What made you start doing this experiment?**

It goes back to early 2000. I was always curious about how some of our wines would do really well some places and not get anything other places. We won the San Francisco International Wine Competition Best of Show. It was a '93 Zinfandel from Pacini Vineyard. And we entered that wine in another competition and it didn't win anything.

I was a judge at the State Fair for a while. I finally told Pooch [G.M. Pucilowski], the chief judge at the time, that I didn't think I was very good at it, so I excused myself. He invited me to be part of the advisory board for the State Fair, partly because of my technical background and partly because I owned a small winery. As soon as I was on the board, I started talking to Pooch about running some tests to check the reliability of the judges.

## **You started testing all the judges in 2005. How did they do?**

The first year, about 10 percent of the judges did pretty well. Another 10 percent did pretty poorly. I would define 'pretty well' [as being] if they could stay within the range of one medal. We could translate the medals to scores. If the judge was within four points, that was really good. If the judge was up to 14 or 16 points disparity, that was really bad.

### **Who were the judges at that time?**

Until this year, all the judges that were invited to judge at the State Fair had to go through a wine-tasting course. If you wanted to be a judge, you had to take a test. There were people who were professors of viticulture and enology at UC Davis and at Fresno. There were wine buyers for stores. There were people who were professionals in other fields, but they were knowledgeable about wine.

This year, we decided to suspend this and let [new chief judge] Mike Dunne invite anyone he wanted to.

### **How did people do this year, compared to the past?**

About the same. The question is whether this program is going to continue. We've been collecting data since 2005. We hoped that we would determine the best judges to use them as mentors.

It turned out that the judges who were really good in 2005 were in the middle of the pack in 2006. That has repeated. The ones that do well are going to change every year.



© AFP/Robert Hodgson | L-R: Panelists tasting wines at the Concours de Vins Mâcon; Robert Hodgson; award winners at the California State Fair

### **Is the California State Fair better or worse than other wine competitions in the U.S.?**

The second paper I wrote had to do with tracking wine through U.S. competitions. About 99 percent of the wines that get gold medals one place, get no award some place else.

Several gold medal-winning wines were entered in five competitions. None of them got five golds. None of them got four golds. It's amazing, the lack of consistency. I put together a study that showed these are the results you would get if this were a completely random process. I'm not willing to bite the bullet and say it's completely random, I don't think that's true. But that's what the results indicate.

One winery entered 14 competitions and got no awards in 13 of them, and got a gold medal in the 14th. Guess what's on the label of that wine? Gold medal winner.

### **Who are these competitions for, the consumer or the wineries?**

They're for neither. They're for the competition organizer. Haven't you figured that out yet? At the [San Francisco] Chronicle competition, you know how many entries there are? 5,500, at \$75 per entry. You can multiply, right? You judged at that, right? Did you get paid anything? You collect 5,500 wines at \$75 and pay the judges nothing. I think somebody should follow the money. I think it would be an interesting article.

The people that enter wines are doing it because they think that having a gold medal on the bottle will sell wine. There's some work that suggests that's not true. I enter competitions because I like to gamble.

**The OIV has developed a system for wine competitions like the Concours Mondial in which tasters get no more than 50 wines a day, and are not supposed to discuss the wines. They give scores to the wines, and their scores are combined and statistically adjusted to adapt for judges' biases. For example, I was on average about 1.2 points below the rest of the judges on my panel this year, so there's an adjustment for that. What do you think of that system?**

What they're doing is exactly right. We've talked about that at the State Fair. You should take into account the bias of each judge. We've said we should evaluate the wine twice. The first time without talking. The second time, when judges talk about it, some judges are very persuasive at getting people to change their score. That data is not very valid.

One of the things we've been working on is using rank-order statistics. One real simple example is, if you have 10 wines [and four judges], you give the wines scores but you convert the scores to overall ranks. You add up all the ranks together. You take the sum of the ranks. Let's say a particular wine got 1st (out of 10), 1st, 3rd, 1st. It would have 6 points. Another wine would be ranked 10th by all four judges, it would have a rank-order score of 40. The wine with the lowest rank-order score wins.

There's a lot of mathematical theory that says that's a better way to do it. Without rank-order scoring, somebody who's a very lenient scorer would have a lot more weight in the group than a more stingy scorer.

### **I'm a stingy scorer, so I have less weight?**

That would be true, but it wouldn't be true if you look at the ranks. I doubt that it's going to happen because averages are easy to understand.



© Phil Kampel | The 17-day California State Fair is held each year in July

**What about social promotion medals, when judges adjust their scores upward because another judge convinces them to?**

That's a human characteristic. That just screws up the data. There's no way to work with the data. If the data aren't independent, you should just give up. I've been on enough panels to know there's a good side to people. They want to help out.

**Do these results apply to ratings given by wine critics?**

These results are only about competitions. I have friends who would like me to submit a bunch of wines to the Wine Spectator with different labels. But I haven't done that. I think there's a real opportunity for the Spectator to enter into an agreement with somebody to test their results statistically. If they do well, they can publicize it and it looks good for them.

I think a lot of the stuff they write about is bullshit. I like to think that I write about something I know about. Clearly people like the Wine Spectator. There's no way anyone can taste that many wines in a week. I don't know how good the Spectator is. All I can tell you is that wine competitions aren't very useful.

**The major wine magazines give one person's opinion, while competitions use a group of tasters. Which should give more repeatable results?**

Definitely a group is better. From a statistics point of view, the more evaluators you have, the less noise. You have to have a panel that has consistent tasters.

What if this were doctors? You'd have one doctor who says, you need to amputate your arm, and another who says, you need to take some aspirin. Would you like a board of experts to determine what woman would be your wife?

**Your results have been available for a few years. Why all the attention now?**

I have no idea, although what you say is true. The last time that there was a fuss was when there was an article in the Wall Street Journal a few years ago. That led to some reaction. And then nothing happened until a month or so ago. All of the sudden The Guardian called, and then the L.A. Times.

**Your research has been used recently to attack not only wine judging, but wine itself. What do you think about that?**

I've only read a few articles. I've only read ones where people have interviewed me and they have been accurate portrayals of what I've said. There are some people who think I'm killing wine competitions. I don't think that's true.

I think wine is a great beverage. I certainly drink wine every day. Most of the time I drink my own wine because I'm cheap. It depends on what I'm eating. One of my favorites right now is a rosato that we made from the dolcetto grape. People still think rosés are inferior wine. That's a shame.