COMPUTING EDUCATION RESEARCH

Raymond Lister

Geek Genes and Bimodal Grades

MANY COMPUTER SCIENTISTS CLAIM that some students are born with an innate ability to program (the "geek gene"), while other students are doomed to struggle with programming. To justify their faith in the geek gene, those computer scientists claim they see a bimodal distribution in CS1 grades. I disagree profoundly with the idea of a "geek gene", so it was a pleasure to read an elegant paper by Robins [4], in which he provides an alternate explanation for why CS1 grades might show a bimodal distribution.

Robins introduces the idea of *learning edge momentum*, which works as follows. Think of a CS1 course as a sequence of topics. Suppose that two students have an equal probability of learning the nth topic but only one of them successfully learns that topic. In Robins' model, the successful student is then a little more likely to learn topic n+1 than the other student is. That is, the student who learnt topic n gains a little momentum, while the student who did not learn topic n loses a little momentum. Robins ran simulations and showed that his model leads to bimodal grade distributions. Since all the simulated students begin with an equal probability of learning the first topic, the simulations demonstrated that a bimodal distribution of class grades is not necessarily evidence for a geek gene – lovely!

Despite my admiration for Robins' paper, my hunch is that the frequency of the bimodal grade distribution is over-stated in the computer science education community. While the bimodal distribution probably does occur sometimes, I question whether it is so common that it is "characteristic" of CS1 grades, as Robins claims.

> I wonder whether grade distributions are less a reflection of the genuine programming ability of students and more a reflection of the methods we use to grade those students.

People tend to see in their data what they expect to see. For example, look at the graph in Figure 1 (page 130) of the famous McCracken et al. paper [3]. The authors of that paper wrote,

> "Figure 1 shows that the distribution of these scores is bi-modal. While the majority of the students did very poorly, there is a second hump in the distribution, indicating a set of students with somewhat better performance."

I don't buy it. I don't see how the small blip high in that graph constitutes a bimodal distribution. I'd classify that graph as a highly skewed normal distribution.

We also need to be wary of edge effects in grade distributions. Students can't score less than zero (a "floor effect"), or more than 100% (a "ceiling effect"), so peaks at either of those extremes are not necessarily evidence of bimodality. For example, take a look at Figure 3 (page 210) in a paper by Blonskis and Dagienė [1]. Some readers might describe that grade distribution as bimodal, but I would describe that grade distribution as being a normal distribution with a pronounced ceiling effect. (The McCracken *et al.* graph shows a floor effect.)

If you think I'm being sophomoric about bimodality, then I recommend that you read an entertaining paper by Schilling *et al.*, "*Is Human Height Bimodal?* [5]" As that paper makes clear, merely eye-balling a graph is a shaky way of establishing bimodality.

While I think the frequency of the bimodal distribution is over-stated in our community, I concede that many CS1 class grades show a skewed normal distribution where most students score low grades, and a few students score high grades (i.e. like the McCracken *et al.* graph). However, such a graph is to be expected in any demanding course with a relatively high failure rate – you don't need to invoke a geek gene, or learning edge momentum, to explain such a distribution.

I wonder whether grade distributions are less a reflection of the genuine programming ability of students and more a reflection of the methods we use to grade those students. For example, Traynor, Bergin, and Gibson [6] provided an illuminating extract from an interview with a student, where the student described his approach to answering coding questions in an exam, when he didn't really know the answer:

> "… you usually get the marks by making the answer look correct. Like, if it's a searching problem, you

> put down a loop and you have an array and an if statement. That usually gets you the marks … not all of them, but definitely a pass".

Dressel [2] described the grades we give students (in any discipline) as

> "an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material."

While I think Dressel may have been overly pessimistic, I also think that most computing academics are overly optimistic about the validity of their grading – so optimistic that it doesn't even occur to them to question the validity of their grading. A good researcher, on the other hand, must consider the validity of their research instrument, so if the computing education research community is going to have a productive discourse about CS1 grade distributions, then we must consider the validity of current approaches to grading. **Ir**

References

[**1**] Blonskis, J. and Dagienė, V. (2008) *Analysis of Students' Developed Programs at the Maturity Exams in Information Technologies*. Third International Conference on Informatics in Secondary Schools. Springer: Lecture Notes in Computer Science, Volume 5090/2008, pp. 204-215.

[**2**] Dressel, P. (1983) *Grades: One More Tilt at the Windmill*, in A. W. Chickering (ed.), Bulletin, Memphis State University, Center for the Study of Higher Education, Memphis, December.

[**3**] McCracken et al. (2001) A multi-national, multi-institutional study of assessment of programming skills of first-year CS students. *SIGCSE Bull*. 33, 4 (Dec. 2001), 125-180.

[**4**] Robins, A. (2010) Learning edge momentum: A new account of outcomes in CS1. *Computer Science Education*, 20(1), 37 – 71.

[**5**] Schilling, M., Watkins, A. and Watkins, W. (2002). Is Human Height Bimodal? *The American Statistician* 56 (3): 223–229. DOI:10.1198/00031300265

[**6**] Traynor, D., Bergin, S. and Gibson, J.P. (2006). *Automated Assessment in CS1*. Proc. Eighth Australasian Computing Education Conference (ACE2006), Hobart, Australia. CRPIT, 52. pp. 223-228. http://crpit.com/confpapers/CRPITV52Traynor.pdf

Raymond Lister
Faculty of Information Technology
University of Technology, Sydney
Broadway NSW 2007 Australia
Raymond@it.uts.edu.au

DOI: 10.1145/1835428.1835434