

# How to Make Benchmarking *Useful*



Scott Beamer  
sbeamer@ucsc.edu

*University of California, Santa Cruz*

- *Ideally* - benchmark / proxy app should be very representative
- ... *thus* innovations/results from benchmark can then directly benefit target application
- *Unfortunately* not guaranteed even if close
- **Representativity matters only insofar as results are useful for target domain**

- Standardize evaluations
- Pick winners
- Track progress over time
- *Guide research by setting target*
- **Not automatic** - need to be pragmatic and critically assess each case

- *Size* - single problem vs. suite
- *Complexity* - simplified vs. full detail
- *Standards* - strict rules vs. loose guidelines
- *Longevity* - single challenge vs. fixed forever
- *Common tradeoff* - simple / flexible may increase adoption but *may* decrease relevance
- **Room for many benchmarks! (multiple goals)**