**Experiment 1**

**Hypothesis:** training judges improves performance of the legal system (i.e time to judgement)

**Design:**

* randomly selected judges get well-funded training in central London venue,
* their judgement decision time will be compared to the national average (control group) over the next two years, if they stay in the Magistrates’ Courts (lowest courts) during that time
* What thre ats to validity do you see?
* What improvements / points to consider would you suggest?

**Experiment 2**

**Hypothesis:** distributing school textbooks improves exam performance in Kenya

**Design:**

* Students are randomly chosen to receive textbooks or not, and their grades are tracked
* Would you randomly select students within schools? Or select some schools to receive books and other not? Why?
* What specific threats to validity do you see? How could they be addressed?

**Praise and criticism**

From a Spectator article:

*When I was 13, my school cricket team received a visit from a top professional cricket coach, an intoxicating visit from the big leagues. I tried to hear what the great man was saying as he watched us, how he advised our teacher. ‘Never praise kids — they only mess it up next time,’ I overheard him say. After pausing to berate me for a below-average cover drive, he whispered to the teacher, ‘It’s different with criticism — that really works.’*

* Is this true? Assuming that it is a conclusion from specific observations,   
  what threat to validity might account for the association in the data?
* Can you find another example for a similar effect at play?