Mathematics Education Research Journal 1993, Vol. 5, No. 1, 28–29

Response to Menon

Glen Rowley
Monash University

This is splendid review paper, well written, forthright and comprehensive. Although the arguments canvassed do not carry us much further than Carver (1978) and a host of other writings, it does bring Carver's review up-to-date, and it addresses the argument to a specific audience (mathematics educators) to whom the matters presented maybe unfamiliar, and it does it particularly well.

Like Carver's article, this one compromises a little more than it delivers. It calls for the abandonment of statistical significance testing (SST), but falls somewhat short of its aims when it comes to what will be done instead. Menon (1993), like Carver (1978), backs off that difficult question, preferring instead to list a few statistical procedures and to conclude that "credible experimental research results can be reported using statistical tools other than SST" (p. 15). And, like Carver, Menon seems to endorse the view that it is better that researchers abandon the use of SST and instead "trust their own informed judgement." (p. 5). I don't really see informed judgement as an alternative to SST, and wonder how, in the absence of SST, informed judgement based on undefined criteria would fare at the hands of critical reviewers like the author of this paper. My guess is that it would fare pretty badly, and probably much worse than SST. Surely informed judgement, supplemented by the knowledge of SST, would be superior to either alone.

It is in this respect that I think the paper falls short. In the absence of SST, how do we reply to the researcher who contends, with the authority of informed judgement, that his or her research has produced conclusive evidence on some question, when with a little computation, it is obvious that the results presented are pretty much what one would expect, even if there were no difference, or no relationship? I would like to point this out, and to advise that if the result is really that important, it is worth investing in another (preferably larger) study with the expectation that it will yield more convincing results. The logic of SST makes this case easy to sustain. Without it the evidence would be equally dodgy, but it would be harder to discount the claims of "informed judgement." I think that the writer is a little inclined to abandon SST on the grounds that it is so frequently misused and misunderstood, without acknowledging that its wise use can contribute to more credible research.

I am happy to see Carver's myths restated, although the myth that $P[H_0 \mid D]$ can be interpreted as $P[D \mid H_0]$ is one that ought to be fairly obvious to mathematics educators, and probably does not need to be laboured. My first reading of Lesnak's last sentence placed the emphasis on the last word—"the probability of this difference occurring by chance *alone*"—if this is so, I think Lesnak may have been referring to $P[D \mid H_0]$ and not $P[H_0 \mid D]$. Nevertheless, the point remains that the difference is both subtle and important, and many researchers have either misunderstood it or been sloppy in their explanations of it. The restatement of Carver's myths is instructive, but the fact that many people misinterpret SST is not necessarily an argument against the use of SST. It may be an argument that SST

should be taught better (and who would deny that it is frequently taught appallingly badly?). If these misunderstandings were corrected, the paper would be shorter, but would the arguments against SST be any weaker?

It is well known that statistical significance is a function of sample size as well as effect size and chosen α (Nunnally, 1964, and Bakan, 1966, demonstrate this better than Hays, 1974). But to acknowledge this is not quite to concede Menon's (1993) claim that it follows that "data can be manipulated to obtain research results which will lead to the rejection H_0" (p. 12). The simple fact is that if your data say $p > .05$ then no manipulation you can do will make $p < .05$. More data may lead to a different conclusion, but that is a different matter. Would it ever be otherwise? I am just concerned that overstatement like this could lead unsophisticated readers to reject SST on totally spurious grounds. Better they read Nunnally or Bakan (or Carver, for that matter) than to gain the false impression that probability levels can be that easily manipulated once you have your data.

My comments should not be taken as reflecting badly on the paper, which I regard as an intelligent, well-argued presentation of an area of dispute that is often misunderstood. If it has a good effect, it will help researchers to understand SST better, and to make informed decisions about its appropriate use.

References

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66,* 423–437.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48,* 378–399.

Hays, W. L. (1974). *Statistics* (2nd ed). New York: Holt, Rinehart & Winston.

Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal, 5*(1), 4–18.

Nunnally, J. C. (1964) *Educational measurement and evaluation.* New York: McGraw-Hill.

Author

Glen Rowley, Faculty of Education, Monash University, Clayton, Victoria 3168, Australia.