

Commentary on *How to Interpret Changes in an Athletic Performance Test*

Christopher J Gore

Sportscience 8, 8-9 (sportsci.org/jour/04/cjg.htm)

Australian Institute of Sport, PO Box 219, Brooklyn Park, South Australia 5032. [Email](#).

[Reprint pdf](#) · [Reprint doc](#)

The summary of the accompanying [article](#) and the PowerPoint slideshow should be essential reading for any sports scientist working with athletes. You don't have to be in such a role very long before a coach will want to know if an individual athlete is increasing or decreasing their score on a particular test compared with their previous test or tests. They might well ask, is an increase of a VO_2max from 5.01 to 5.05 L/min "real"? To answer them it means that you need to know how much an elite athlete is likely to change over a finite period and how much noise is associated with the measure. This work by Will Hopkins combines elements from several of his previous publications, improves and simplifies them, and now provides a clear pathway to answer a concerned coach or athlete.

The slideshow starts with a simplified account of how the variation in performance of individual elite athletes in competition gives rise to a "worthwhile" change: an enhancement that increases medal-winning prospects of one of the athletes in a well-matched group. Novel in this slideshow is Hopkins attempt to quantify a worthwhile change for a team sport athlete. He has chosen 0.2 of the between subject standard deviation (SD). This is a useful start point but, as he indicates, there is no known relationship between fitness test performance and team performance. For instance, teleologically, it makes sense that fast sprint speed and high aerobic power would be advantageous in a team sport such as soccer, but the ball handling and game-reading skills also come into play with the team results. Furthermore, at higher levels of competition a group may be more homogeneous and thus the between subject standard deviation will also be reduced. Nevertheless, it follows that worthwhile increments are also smaller as athletes rise toward the top of any measure.

Hopkins reminds us of the importance of both test validity and reliability, and that reliability is paramount. In Australia, we have been working for more than 10 years, with the geographically remote state sport institutes, to quantify test-retest reliability as a means to understand laboratory and field physiology tests, such VO_2max and 20-m sprint times (Gore, 2000). First we worked on test reliability and after a number of years we moved toward test accuracy, whereby as much equipment as possible is calibrated against first principles of time, distance and mass. Incorrectly, we used total error of measurement to quantify our reliability, but have subsequently used typical error (Hopkins, 2000) to quantify test-retest error and found little difference, owing to small changes in the mean.

Hopkins suggests that you can use published studies to identify reliable tests that you may wish to use for athlete testing. Our experience in Australia in the field of exercise physiology suggests that it is essential that you establish your own typical error using your own athletes and own equipment. It is poor science to rely on others and assume that your error is as low as theirs. You owe it to your athletes and coach to quantify the likely error of a given test in your hands. This can be achieved readily by conducting a test-retest a few days apart on your athletes in a specific squad. Hopkins notes that longer periods between tests, when athletes begin to show individual changes in fitness, are appropriate in the context of interventions of similar duration.

Hopkins recommends using likely limits as a suitable method to provide feedback to coaches and athletes. In Australia our state sports institutes have adopted the "rules" approach as being most expedient. We have also been conservative and sometimes interpreted that useful changes are at least greater than ' $\sqrt{2}$ x noise', which means at worst we are right more than 62% of the time. Contrary to Hopkins' advice we even use a 95% level of confidence when using skinfolds (Woolford and Gore, 2004). This measure is not really a performance test, but thoughtless interpretation can have profound consequences with athlete body image and even eating habits. Thus, in this rare case, I believe that such a conservative approach is warranted.

Hopkins summarizes that you should be up front about the noise when you feed back the test results to an athlete or coach. All reports of physiological tests issued to athletes by our state sport institutes follow that format with the test-specific Typical Error included and a note in the footer explaining the rules for interpretation.

Overall, I believe that anyone working with small groups of athletes is flying blind if they don't know the typical error of the tests they are using. Interpreting meaningful or worthwhile changes in test results has been considered a bit of an art in some circles, but the science of Hopkins' approach allows one to be confident about the degree of uncertainty of their recommendations.

Gore C (2000). Quality assurance in exercise physiology laboratories. In: Gore CJ, editor.

Physiological Tests for Elite Athletes. Champaign, IL: Human Kinetics, pp 3-11.

Hopkins WG (2000). Measures of reliability in sports medicine and science. Sports Medicine 30, 1-15.

Woolford SM, Gore CJ (2004). Interpreting skinfold sums. Use of absolute or relative typical error? American Journal of Human Biology 16, 87-90.

[Back to article/homepage](#)

Published Nov 2004

©2004