the CrossFit JOURNAL ARTICLES

What is Meaningful

Lon Kilgore, Ph.D.

To the objective observer, it should be fairly obvious that CrossFit methods of fitness training are proving themselves in the field. Out in the real world, the average Joe who sees results like those typically seen by their CrossFitting friends are swayed by success. This is why the CrossFit community is growing and thriving. But there is always a cadre of exercise scientists and physicians who don't necessarily believe results from the field (after all, "there were no controls"). There is an adage in the sciences that "you can prove anything with a single case example," so anecdotal reports of success from the field are frequently assigned a merit and validity best suited for File 13 or Area 51. If the testing didn't happen in a controlled laboratory environment, the thinking goes, the results cannot be the product of an evidence-based system and therefore must be the worst kind of popular and faddish trash or fiction.

But does it really matter what exercise scientists say? The disregard some academics have for practitioners is a two-way street. Most exercise scientists know that the research reports or theoretical papers they publish are completely ignored by actual practitioners. In a very recent conference keynote speech, Dr. William Kraemer, putatively one of the most recognizable and respected figures in exercise research, said "Coaches don't listen to sports scientists." If such a lofty scientist expects his research to be ignored, what hope is there for the rest of us researchers? Not much, I'm afraid. It is a frequent reality that people in the field find results from the field more meaningful than any results from any exercise science laboratory, and often rightfully so. That's a pretty damning statement coming from a sports scientist, but let's use a quick example from one of my primary areas of scholarly interest, muscle hypertrophy, to make the point as to why this is the case.

The average competitive bodybuilder has a muscle mass about 75 percent greater than the average Joe. Yes, some of the competitors are benefactors of chemical enhancement, but there are those who have dedicated their lives to getting big and have done so clean. So, in the field, a halfway decent coach who got half of the average results seen in competitive bodybuilders might see around a 35 to 37 percent increase in muscle mass in an athlete who trains religiously without exogenous hormonal enhancement. That's not a stellar result, but compared to no training or really bad training, it is a large improvement. Now let's compare the results that are normally seen in the best exercise science laboratory studies to the results seen in the field. A long-term laboratory study that has a large significant result will demonstrate at most a 20 to 25 percent increase in mass (most produce low single-digit results). This is pretty slim compared to the gains seen in real-world competitors, or for that matter compared to the results of our hypothetical halfway decent coach. How does this happen? These studies use the most responsive group of trainees possible, beginners, so big changes should happen very fast. But, alas, there is a large contingent of exercise scientists and physicians who believe that one work set of an exercise is enough to achieve maximal results. If they do decide to use multiple sets, many researchers think that three sets of ten will produce the same results as ten sets of three—and they will use the same weight for both organizations, a weight described as "low to moderate intensity." But as is obvious to anyone who actually trains "in the field," very little stress causes very little adaptation, regardless of the set and repetition scheme. So training methods used in the lab are generally substandard, and conclusions based on those methods really do not have much relevance to making people more fit. It seems as though most exercise researchers misapply the basic tenets of human adaptation.

But missing the point of how the body adapts to a stressor is not the only place where exercise scientists go wrong. In a recent article in a well-known exercise physiology publication, a faculty member from a school of physical therapy made an astonishing conclusion about training and competition. He proposed that a single set of five to eight repetitions with a moderate weight is appropriate as a warm-up prior to a training session or a weightlifting/powerlifting competition. Any weightlifter or powerlifter, from rank novice to world-class elite, will tell you that a traditional multi-set and low repetition warm-up is needed to prepare the body to neurally and efficiently handle maximal efforts.

It's not just the scientists on the strength side that seem to have a problem re-creating the real world in their laboratories and actually generating useful information. If we consider the concept of VO_2 max, the soul of aerobic exercise physiology interests and dogma, we find that a large number of exercise scientists believe that VO_2 max, is only minimally trainable. They propose maybe only a 5 to 10 percent improvement with training as the limit of possibility. That would mean that someone like me, with a 48 ml/kg/min VO_2 max, would never be able to compete at any event requiring more than 53 ml/kg/min and have a hope of being competitive. But again, let's look at the real world of high-level competitive athletics and a familiar figure, Lance Armstrong. In Armstrong's early days of high-level competition (as a triathlete), his VO_2 max was in the low 60s. During his astounding run as Tour de France champion in the past seven years, his VO_2 max was reportedly in the upper 80s. This is a bit more than the 5 to 10 percent considered to be the top end of the improvement spectrum. OK, there is always the possibility that some special assistance of exogenous origin was involved in this example, but the point is that this magnitude of improvement is not uncommon in aerobic athletes who train progressively and appropriately under the astute eyes of expert coaches. So, here we have another example of laboratory models of training resulting in very small single-digit percent gains—far less than the gains of up to 33 percent generated by current field training methods. There apparently is a dark and deep chasm existing between the real world of training for peak performance and the ivory tower of exercise and sport science.

The lack of relevance stems, in part, from the lack of useful field experience on the part of many exercise scientists. It is not enough to have an interest in and theoretical familiarity with a subject. To make exercise science research useful, there must be a solid foundation in practical application. If you are going to do research in an area, you just might want to have a good deal of participation and/or coaching history in what you propose to research. A good experiment in exercise science begins with observing results from the field and then asking why and how things work. What happened to your body or your trainee's body during training that generated the results of interest? Isn't that what a scientist is supposed to do—observe, question, experiment, analyze, modify, and repeat? If I can understand why something that I see happening in the field occurs, either from observation and explanation or from observation and then experimentation, I can help other people make wise choices in exercise program design. If I fail to derive and frame my initial observations and my experimental design in practical application, then everything downstream suffers. This is much harder to do than it sounds, so while we can be critical, we needn't be abusive to our scientific brethren; they really are working within a flawed system that almost enforces the production of limited-utility research.

Sport and exercise science is, after all, an applied science. The implication of that is that we are supposedly trying to solve specific problems—problems related to improving fitness. Sport and exercise science is not a basic science, and its professoriate is not supposed to be intent on creating knowledge for knowledge's sake. A physiologist can do basic science because he wants to know how something works for no other reason than just to understand it. This is a noble endeavor and, in fact, the last time I checked, all Nobel prizes in the sciences, except for one, have been awarded to basic scientists. That's OK but an exercise or sport physiologist is supposed to do applied science for the reason of solving a problem or knowing how to make something—the human body, in this case—work better. We don't need to randomly manipulate some obscure variable just because it's never been done before. But this is precisely what happens in the academic world of exercise science, physical education, kinesiology, human movement sciences, or whatever other name we go by. In the race to get published for tenure and promotion, in the race to produce a thesis or dissertation, we are all rewarded for doing something no one else has done before. This leads to some less-than-useful research done in the name of originality—or novelty.

It hasn't always been this way. In the 1940s and before, exercise research and teaching was firmly grounded and of high quality. Research was done in the name of military readiness or work physiology or efficiency and was conducted in biology departments and medical schools around the world. But things changed. The bastardization of scholarship that came from putting varsity athletics in bed with academics produced a set of academic units with internally competing and divergent goals. The end result was a rapid descent of exercise science from being part of mainstream biological science to the verge of being designated a non-academic endeavor due to the lack of a significant and unique scholarly body of knowledge specific to its new bedfellow, physical education.

The Fisher Act of 1961 in California mandated that any university major must have as its underpinning a unique body of evidence supporting practice. In the California state legislature's opinion at the time, to coach, to teach PE, or to train someone required no special knowledge or skill, since, after all, if it did, there would be a body of knowledge to guide practitioners. The major exercise professional organizations and academic units went into overdrive to develop a research database to support the discipline's existence (and no, all the old military and work physiology stuff apparently was not considered "unique" to "physical education"; although it was very relevant, its science was labeled as originating in the disciplines of biology and medicine rather than constituting one of their own). Disciplinary journals were created to publish reports from physical education scholars, exercise physiology labs, biomechanics labs, sport psychology labs, and any other newly created exercise-related lab that did a version of science. The rush to publish the evidence to support practice set a precedent for less than stringent editorial process to facilitate the rapid creation of a database of literature that would justify physical education departments and keep them on campuses.

Unfortunately, that process is still lax in many exercise journals over 40 years later. Old habits are hard to break. And funding of research contributes to the problem, in that funding agencies don't back performance research intended to establish the route to maximal fitness. Instead, they fund health research that attempts to determine the minimum amount of exercise needed to stave off disease and death. This is a fundamental and catastrophic error on the part of government funding agencies and means that anyone interested in developing training modalities for making better athletes, stronger workers, or fitter citizens through research was, and is, out of funding luck. It also means that truly meaningful exercise information will continue to be scarce.

But do you need to be on a university campus and hold a Ph.D. to be a practicing exercise scientist? Actually, no. Greg Glassman, Louie Simmons, and Mark Rippetoe are all great examples of "exercise scientists" who work outside academia. Their gyms are their labs and their athletes are their research subjects. Scientific method applied to relevant programs and real trainees is how all of these eminent coaches arrived at their present training methods. Although the methods may be "unscientific" and the analyses simple (it works or it doesn't work), scientific method applied to exercise resulted in three very relevant and successful training methods (CrossFit, Westside, and *Starting Strength*, respectively). In your practice or in your own training you can do the exact same thing: observe, experiment, and adapt.

I doubt that Glassman, Simmons, and Rippetoe ever did any in-depth statistical evaluations of their training results beyond arithmetic calculations of improvement. University exercise scientists, however, are mired in statistics. We like to use them to estimate the worth of our research and assess whether the findings might be reproducible. We use the term "statistically significant" to describe any finding that departs mathematically from a reference standard. The problem is: does a "statistically significant" assessment mean that a finding is useful or, perhaps more to the point, does a statistically insignificant finding mean a lack of merit? Consider the idea of a "p-value", the probability that a finding would be replicated if the experiment were performed again. We always select a p value of 0.05. This number represents the probability that repeating the experiment we would have similar results 95% of the time. The concept is a pretty good one—adopt a rigorous and mathematically objective standard in order to determine whether or not to accept the results of the study.

However, a quite learned maverick strength coach, Loren McVey, once pointed out to me a flaw with this black and white approach to determining whether a finding is meaningful. He had read three very similar papers and each reported "non-significant" increases in performance after a type of training, as they found p-values higher than the 0.05 set as the standard for significance. Each of the three papers reported p-values of about 0.10 or about a 90 percent chance the result were reproducible and real. Statistically, and therefore scientifically, the training protocols used were not deemed useful, because they needed a 95% probability. Statistics aside, though, here we have three separate papers all finding similar improvements in performance of similar magnitude from similar protocols. Although the findings were replicated separately, none of the individual studies were strong enough to make a statement supporting the use of the training protocol investigated. But can we overrule the dismissal of the individual studies by considering the three papers collectively and arrive at a more utilitarian and combined statement that just maybe the training approach presented in the three papers has merit and that a 90 percent likelihood of replicating success is pretty good? Maybe so. The coach or trainee must make that decision, and if you are an educated and independent thinker you have every right to interpret data in a manner consistent with your decision-making standards.

A final consideration on knowing what is meaningful: coaches and trainees must read and must think about what they read. And I'm not talking about the typical bookstore muscle and fitness mags, you must read high quality textbooks and experimental reports in scientific journals. OK, some of the journals are less than ideal (refer back to Rippetoe's article "Conventional Wisdom and the Fitness Industry" in *CrossFit Journal* issue 54), but, just like learning from a bad coach, you can learn something from bad journals. The least you can learn is what doesn't work. That may be the opposite of what most authors intend for their papers, but learning is always meaningful no matter how it is accomplished. This particular requirement is not commonly practiced by but a small minority of typical strength or fitness practitioners. But then again, CrossFit instructors and CrossFitters are not so typical. It is evident from reading many of the daily trainer and trainee posts on the CrossFit.com web page that this community thinks and reads more broadly than the average fitness professional and fitness trainee. By focusing some reading time on exercise science, maybe, just maybe, you'll be like Loren McVey and happen across a number of similar studies that together provide a unique and practical insight that will help you turn out a more fit human. But more likely you'll be appalled at the lack of usefulness of the research report you have in your hands and you will wave it around and make every one of your peers read it and then you'll all discuss it and bash it about in animated conversation. Guess what? That's good. That's meaningful. You've read, you've analyzed, you've discussed, you've made an educated judgement of merit, and you've furthered your personal definitions of the boundaries between utility and futility. When all things are considered, this is important because, at the end of the day, it is your judgment, your ability to logically defend your theory of and approach to training, and, ultimately, your ability to produce training results that is truly meaningful.

◇

Lon Kilgore, Ph.D., is professor of kinesiology at Midwestern State University, where he teaches exercise physiology and anatomy. He has extensive experience as a weightlifter himself, and he has worked as coach and sports science consultant with athletes from rank novices to collegiate athletes, professionals, and Olympians. In addition to publishing articles in numerous scholarly journals, he is co-author, with Mark Rippetoe, of the books *Starting Strength: Basic Barbell Training* and *Practical Programming for Strength Training*.