
THE CrossFit JOURNAL

Testing Fitness as Sport

The CrossFit Games are the ultimate test of fitness.

Tony Budding shares some of the theory behind the structure of the 2010 Games.

By Tony Budding

September 2010



A. Anthony/CrossFit

CrossFit is both a training modality and a sport. As a training modality, we can improve the real-world physical capacity of folks from any walk of life. As a sport, we can compete for fun, or we can compete for the title of “fittest.” The CrossFit Games are the world championships of our sport, so we title the winners the Fittest on Earth. Using a single CrossFit workout as a sport is pretty simple. Put the names up on the whiteboard. Allow scaling, which is a form of handicapping, or not. Pre-register any excuses, or not. Set the terms of the workout and go. The winner is he or she who finishes first, with the most rounds and reps, or who lifts the most weight.

Combining multiple workouts into an integrated CrossFit competition is more complicated. Because every workout is different, combining them in a fair way is rarely a simple, straightforward process. We've been experimenting with this in the Games for four years now. The rest of this article is a philosophical look into what it means to test fitness as a sport.

What Is Fitness?

Testing fitness is not easy. The technical definition of fitness is "increased work capacity across broad time and modal domains." A more casual definition of fitness is "competency at the tasks of life." Both of these definitions are inherently hard to pin down. This is not a flaw in the definition but rather a complexity inherent in what fitness really is. In other words, artificially simplifying either the definition or assessment of fitness doesn't change the complex nature of fitness.

The tasks of life are incredibly diverse, probably even infinitely diverse. There is no possible way we could assess capability in all of them in a year, not to mention a weekend. Therefore, we must seek exercises—alone and in combination—that both represent and predict broad capacity.

Obviously, functional movements must be the foundation of any good test of fitness.

Functional Movements

Obviously, functional movements must be the foundation of any good test of fitness. Functional movements are the natural, prehistoric and essential movements of life. They are characterized by their ability to move large loads long distances quickly. They are the best tools for delivering the highest possible levels of average power (real work performed divided by time of completion). There is simply no way to assess fitness without pushing the limits of power in varied domains.



D. Re/CrossFit

To mimic the real world, the CrossFit Games asked athletes to go overhead immediately after a longer effort.

What, then, are the relevant domains that best assess and predict fitness in other domains? There is no simple answer to this, and it is here that the richest, most productive debates about fitness occur. Because life requires millions of specific tasks, we have to be satisfied with demonstrating capacity in categories or genres of movements. Some functional movements require very little specific skill; others require much more. The fittest should be competent in both. Some tasks require you to control your body in three-dimensional space (gymnastics, calisthenics, running). Others require you to move external objects (weightlifting, throwing). Some movements are performed standing. Others require inversion (handstand push-ups) or getting horizontal (burpees). Some external objects start on the ground and must be lifted. Others must be moved some distance. Some objects are easy to grab (barbells, kettlebells) while others are less so (sandbags, stones, tires). Sometimes external obstacles have to be navigated (over, under, around).

A. Anthony/CrossFit



Odd objects made an appearance at the Home Depot Center.

Combinations, Not Single Modality

In the 2010 CrossFit Games, we included all of the above. We also tested them in combination and at various loads and volumes. It is becoming increasingly clear that single-modality events are inferior tests of fitness when compared to well-designed multi-element events. Single-modality days are essential training tools, but given the inherent limitations of any test, well-formed combinations of movements are much better predictors of capacity across broad time and modal domains. Even workouts like the final event in the '08 Games—30 reps, ground to overhead, with 155 lb.—aren't as good at assessing and predicting fitness as the same drill combined with something very different (such as gymnastics or calisthenics movements).

The reason for this should be obvious. Single modalities are less broad, and thus it's possible an athlete could have a narrow capacity that just happens to correlate to this particular event. The likelihood of an athlete having two narrow bands of capacity that are both tested in a single event is tiny in comparison. The challenge of course is creating combinations that achieve the desired impact.

Selection of Movements

There have been complaints that the Games each year are more a test of the best CrossFitter than a test of the world's fittest athlete, with the evidence being the high correlation of exercises in both CrossFit workouts and Games events. These arguments tend to suggest we should get rid of all gym-style implements (barbells, dumbbells, etc.). This is based on a misunderstanding of why we use gym-style implements in the first place. Very simply, they are unmatched in their ability to test (produce) power output in extremely varied manners. Fitness is work capacity (power output) across broad time and modal domains. How better to test capacity at heavy, medium and light loads with various durations and volume than with an implement that is ideal for maximizing power?

Thrusters at 45, 95 and 185 lb. are testing different capacities. We know this because the rankings of the same athletes can change significantly when the loads and volume differ when keeping the movement constant. Furthermore, odd objects present substantial logistical challenges. Consistency across all objects, efficiency of testing, varying time and modal domains, and visibility for spectators are all more difficult with odd objects.

It is becoming increasingly clear that single-modality events are inferior tests of fitness when compared to well-designed multi-element events.

Obviously, we still use odd objects despite these challenges, as we used sandbags, wheelbarrow and walls of various kinds and sizes in the 2010 Games.

One final limitation of odd objects is that they tend to tax the same aspects of the body. Testing variation of capacity is thus more difficult if you were to use odd objects alone.

The Events

The nine events of the 2010 CrossFit Games covered a wide variety of skills and domains. The athletes' outputs at each element were compared to those of their peers, and the most consistent top performers made it to the podium. What follows is a breakdown of the various elements in each event and why they were chosen.

The first event of the Games was 9, 7 and 5 reps of muscle-ups and squat snatches (135/95 lb.). This was a relatively low-volume, high-skill, high-power event. We knew everyone could do both muscle-ups and 135/95 squat snatches, but we also knew the couplet required enough skill and had enough volume and weight that athletes would be able to differentiate themselves nicely. We required a single-movement squat snatch, something new in CrossFit competition, because it required greater amounts of coordination, accuracy,

agility and balance. The combination required a very well-rounded athlete; testing either just muscle-ups or just snatches would have delivered a different ranking.

The second event was the long Helen-like workout, followed immediately by a max-effort overhead lift. We chose the long Helen because it was fundamentally opposite to the first event. It was long, light and high volume. None of the movements were complex, new or particularly challenging to Games competitors. The running distances were long enough to punish inefficient runners but not so long that they required specialization. The 72 pull-ups were a lot but not an excessive amount because workouts like Angie and Cindy require 100-plus pull-ups in a similar time frame.

The athletes had 90 seconds after completing this first part of the workout to get a single maximum load overhead. We did this for several reasons. The first is that testing overhead capacity at high heart rates is a new endeavor for competition, but not life. We regularly hear from soldiers and first responders that max efforts on duty (lifting, fighting, etc.) often follow intense sprints or



S.Dy/CrossFit

The high-skilled, high-power movements of the first workout created a challenge for all the athletes, including 2008 Games champ Jason Khalipa, shown here missing a squat snatch in dramatic fashion.



S.Dy/CrossFit

Inversion with a twist.

other intense efforts. Furthermore, the 90-second time constraint required that the athletes demonstrate significant self-awareness in knowing how much they could handle. There simply wasn't a lot of time to experiment. This is also similar to many situations in the real world.

The third event was a medium-duration, moderate-volume, mixed-element workout. It was a seven-minute AMRAP of deadlifts (315/205 lb.), pistols and double-unders, with short sprints between the deadlifts and the pistols/double-unders. The heavy deadlifts and runs were relatively low skill, but the pistols and double-unders were relatively high skill. This workout required you to be strong and nimble with lots of gas. Any one of the three elements could trip you up if it was a weakness.

The fourth event was moving sandbags from one part of the stands to another, navigating stairs, walls and wheelbarrows along the way. There were four different sizes and weights of sandbags spread out. You had to strategize the order in which you'd do things, but the rules were very simple: move the bags. Were you going to make fewer heavier trips or more lighter trips? How well did you load the wheelbarrow? Most of the biggest differences in performance came from avoidable situations, such as tipping over the wheelbarrow and not shifting unsuccessful strategies (i.e., trying the same technique over and over even though it wasn't working). Adaptation, planning and staying within limits are also elements of real-world fitness.

The fifth event comprised heavy cleans and handstand push-ups. CrossFitters should be proficient at both. In planning this event, we went back and forth quite a bit. It needed to be new, hard to complete, fair to the competitors, easy to judge and fun to watch. When we pulled the trigger on the format, we assessed this workout to be fairly balanced. We thought an elite CrossFitter with no significant chinks would struggle equally with the handstand push-ups and heavy cleans. In the end, though, most athletes struggled much more with the handstand push-ups than the cleans.

I've wondered about this. Was the workout unbalanced or were these athletes generally weaker at handstand push-ups than heavy cleans? Ring handstand push-ups are hard, but because competitors were allowed to wrap their feet, they were not an entirely different skill than regular handstand push-ups. In other words, the athletes who were best at regular or parallette handstand push-ups should also be best at the ring variety. Similarly, for the women, the slight extra depth ended up differentiating the proficient from the barely adequate.

One interesting observation here is that there has been a vocal emphasis among the general affiliate community on "strength bias," referring primarily to the ability to move heavy external objects. This type of strength is certainly important, but fitness also requires the strength to control and move your own body. Four handstand push-ups is not a huge number, even on the rings or plates. The relative ease of the heavy cleans and relative difficulty of the inverted push-ups for so many of these athletes appears to me like a training imbalance.

The Final Event

The sixth event was the blind triple event. We took simple movements that all the athletes would be familiar with and combined them in ways that would pack a punch but still be reasonable. We had a high enough volume that most athletes wouldn't be able to finish, which in essence turned this into a single long event. The blind element meant little to no opportunity to plan or "game" the workout. It also meant the athletes didn't know how long they were going to have to work. So they had to go as hard as they could while still leaving something in the tank for the next event(s).

The tests needed to be both hard and fair, where athletes could differentiate themselves through their performances.

This again mimics the real world, where you rarely know how long something will last or what comes after. The logistics of pulling this off were complex to say the least. We had to keep all the events secret from all the athletes until they were ready to move. If anyone knew ahead of time, it would have dramatically altered the unknown aspect of the competition. We had to have movement standards and combinations that could be easily described and judged and quickly understood by tired athletes. And, the tests needed to be both hard and fair, where athletes could differentiate themselves through their performances.

All three segments were scored separately, which did several things. It recognized the difference between starting strong and finishing strong. It also allowed for greater movement of standings if an athlete dominated or if we exposed weaknesses. Also inherent in the scoring was that every rep counted. The scoring approach of the Games was to rank each athlete compared to his or her peers. Time caps are arbitrary limitations. Folks who complete the workout faster rank higher than the slower

athletes. Athletes who complete more reps (though not the entire event) rank higher than those who complete fewer, but behind everyone who finished them all. This was consistent and fair for all timed events.

Moving into the specifics of the event, the first part was push-ups and overhead squats separated by a 12-foot wall: basic calisthenics, climbing and relatively high-volume overhead barbell work. The second part was toes-to-bar and ground-to-overhead: relatively high-volume body control and barbell control, both with a substantial systemic hit. The final part was wall burpees and rope climbs. How well can you move your body down, up and over? All three of these were fairly simple, moderate-weight drills that tested the athlete's ability to keep going. How much do you have in the tank? How deep can you dig?



D. Smith/CrossFit

The final event of the Games found athletes climbing walls and ropes for time.



S.D./CrossFit

Kristan Clever climbs the 20-foot rope en route to victory in the Games.

The Rope

There have been some complaints about the lack of matting under the ropes. Most of the complainers have gone so far as to express incredulity over its absence, as if having mats in place was indisputably the proper procedure. This is an oversimplified and naïve position. While it's not an experiment I'd like to try, I'd practically guarantee that having gymnastics mats under the ropes would have dramatically increased the number of injuries, not reduced it.

Of course, there is an important place for matting. For example, if someone were to freefall from any distance, I'd hope there was matting under him or her. But, even better, would be to avoid the freefall in the first place. The assertion that not having matting was irresponsible misses the very important point that the lack of matting resulted in more conservative choices on the part of the athletes. In other words, freefall was all but eliminated in that third segment.

Before we get into all the "what-ifs," we have to be clear that there were only two minor injuries that actually happened, and this includes all the individual athletes plus the teams (1). I'm confident that fear was a significant factor in preventing additional injuries. Without this fear of falling, which would have been less severe with any kind of matting, the athletes would have taken more risks. Several athletes came within a few feet of the top and decided to come down because they didn't think it was safe to keep going. This is called intelligence, which is a part of fitness. If there had been mats, some athletes would have felt safe enough to keep going. This would have been stupid. A fall from 18 or 20 feet is very dangerous even if you know proper falling techniques. Furthermore, mats are inherently unstable, and falls from even a few feet could result in badly sprained ankles or significant damage to knees and hips.

In other words, while good gymnastics matting might have helped reduce injury in a theoretical catastrophic fall, it may have caused one by giving an athlete a false

sense of security. Furthermore, a much more likely scenario is that with matting present, athletes would have been very likely to drop from significant heights to gain competitive advantage. But landing safely on matting is difficult for trained athletes and extremely risky for untrained athletes.

Scoring

The purpose of the CrossFit Games scoring system is to determine the Fittest on Earth. Before the Games, there was an extensive qualification process by which about 100 men and women earned the right to compete in the arena at the Home Depot Center. Just getting there established them among the world's elite of fitness. The Games just needed to differentiate among these elite.

After last year's Games, there was a lot of talk about proportional scoring. The great benefit of this system is that it rewards margins. If I beat you in Elizabeth by one second but you beat me in Diane by three minutes, we can say you're fitter because of the margins of victory even though we each won an event. This impact of marginal differences, however, becomes less significant as the number of events increases. But the fatal flaw of proportional scoring is that the margins and proportions between different events are not equally valid indicators of fitness. Some workouts simply have greater margins, even as a percentage. In fact, the workouts with greater margins usually have specialized skills in them. In other words, when you dig into the reality of proportional scoring, it favors the specialist by overly weighting workouts with special skills.

At the end of the day, no one was taken out of contention because of some mathematical anomaly. The best athletes differentiated themselves early and often.



T. Campbell/CrossFit

Graham Holmberg didn't finish below 16th in any event at the Games and came away with the overall title.

This year, the most common complaint about point-per-position scoring has been the impact of cuts to the field. It is mathematically true that if you finish at the bottom of the heap before the cut, the cost is greater. If you finish 40th in an early event but still make it past the first cut, the worst your competition can do is 24th. In other words, a bottom finish in the first four events leaves you with 40-plus points, but after the cut the most points you can get is 24 or 16. This makes deficits harder to eliminate as the competition goes on. The complaint is that this means the early events are weighted more heavily.

This is not quite true because there is a better way to describe this mathematical impact. It's not so much that it weights the early events more heavily but rather that it

punishes glaring weaknesses. This system intentionally rewards athletes who finish closer to the top in every event. This assumes, of course, that the early events are good tests of general fitness. Furthermore, the number of athletes in the initial pool is a huge component. This year 45 men and 41 women competed. They all qualified through a solid process, but were they all legitimate contenders for the title? Of course not. Out of the bottom 20 of each group, there were only four top-10 finishes in any of the first four events. Even with all the variety built in, they were never in contention. Eliminating them at the first cut, therefore, didn't really change the competition.

Going deeper into the actual results, neither Graham Holmberg, Rich Froning Jr. nor Mikko Salo finished any event below 16th. None of them was punished by having a larger field in the early events. Chris Spealler had a 26th and a 22nd. If you capped the score on any workout at 16 (the fewest number of competitors in any event), he would have finished with 16 fewer points. But he was 18 points behind Rich, so the overall placings would have been identical. Even Austin Malleolo, who got 37 points in the max-overhead event but still finished sixth overall, wouldn't have made the podium with a cap of 16 points because Speal would have benefitted from that also.

Going deeper, only three men in the top 16 after four events (right before the cut) had over 30 points on any event. On the women's side, no woman in the top eight overall had over 20 points in any event, and no one in the top 16 after four events had over 30 points on any event. What does this mean? It means that the scoring system worked beautifully.

At the end of the day, no one was taken out of contention because of some mathematical anomaly. The best athletes differentiated themselves early and often. The system rewarded those athletes with the greatest work capacity across broad time and modal domains. Whatever absolute mathematical facts there may be about weighting early or later events more or less, the reality is that the Games, like nature, favor the generalist and punish the specialist with any glaring weakness.

The bottom 20 athletes were never really in contention, so the mathematical weighting of the early rounds was insignificant. Even when the pool of athletes was reduced to 24 and then 16, there was very little change in the top positions. This means that the best athletes performed consistently toward the top in spite of

how large the field was or what the events were. This is particularly impressive because of the diversity of athletes, movements and events.

A True Test of Fitness

Each year, we refine the CrossFit Games, and each year they become a better test of fitness and a better test of the world's best athletes. In 2010, the events were more balanced, more varied, and tested the athletes' work capacity across broader time and modal domains. With the point-per-position scoring system, the truly elite of these awesome competitors differentiated themselves from their peers by finishing consistently toward the top despite the variety.

In this way, the programming and structure of the competition matched the scoring system. With fewer scoring events, or more specialized events, or with different cuts, the competition wouldn't have been as fair. These elements worked together to ensure that the athletes who finished on the podium at the Games were indeed the Fittest on Earth.

Footnotes

1. Rich Froning Jr. bruised his heel when he fell, and Heather Bergeron had a minor sprained ankle when she landed on the rope.



About the Author

Tony Budding is the Director of Media and Web Content for CrossFit, the Co-Director of the CrossFit Games, and the Executive Producer of Live Media for CrossFit.

