Norms vs. Criterion-Referenced Standards

In the past, many fitness tests had test users interpret test scores by comparing them to norms. *Norms* are created by gathering test results from a representative number of people from a large group. For example, if you were going to compare children in your 5th grade class to children across the country, the test results would be collected from children at that grade level in all 50 states. When an individual child's test score is compared to this group, usually the result is expressed as a percentile. For instance, if a child scored at the 70th percentile, 30% of those who were tested would have performed better and 70% worse.

Norms have some inherent flaws. First, for the percentile to have some meaning, you must be sure that the child is being compared to norms that were established using the right reference group. Norms for teens, for example, wouldn't be a good comparison for a fourth-grader. Neither would norms developed from test results in only one region of the country if the comparison was supposed to reflect children across America.

Second, norms are meant to be used for comparisons of individuals, rather than groups. The norms are based on individual results, and they tend to give group averages a low estimate if the results are higher than the middle of the distribution and a high estimate if the results are lower than the middle.

Setting a particular percentile as a standard has several disadvantages. It compares students to the current level of performance of other similar children, rather than to a level they ought to achieve for good health. If the standard is high, it also may discourage those with low or moderate fitness levels. Since genetics is a component in determining fitness levels, some students may have difficulty reaching a high standard even if they put in a good effort.

More useful for promoting health are *criterion-referenced standards*. Such standards are based on research, tying a level of fitness to a good health outcome. To have a criterion-referenced standard, a fitness test must measure some element that is part of being healthy. The PACER test is a good example, as the aerobic capacity it measures can be related to health outcomes such as heart health.

While criterion-referenced standards are more closely related to desirable health outcomes than norm standards, they also have some flaws. First, the accuracy of the standards heavily depends on whether enough appropriate research has been done to establish them. Second, the standards may be low enough that very fit children can meet them too easily and perhaps not be motivated to try as hard to keep or improve their fitness. Finally, the validity and reliability of the standards must be determined through studies as well, to make sure the test results are truly correlated with the intended outcome.