The evaluation of phantom image quality depends upon a number of variables such as viewing conditions (ambient light illuminance, viewbox luminance, masking), the phantom film optical density, variability among phantoms, and variability among readers. An evaluation of 235 Mammography Quality Standards Act (MQSA) inspectors in 1997 was performed to address the issue of consistency of their phantom image quality scoring. A set of 5 different phantom images were duplicated and scored by each inspector. The inspectors each received an individual test packet consisting of the 5 films, with each film labeled differently, and were instructed to score the films using two different criteria. Each individual scoring sheet and film packet was returned. The primary difference between the two criteria was that one was independent of artifact subtraction, the other included artifact subtraction. Consistency was measured as the coefficient of variation for each of the sixteen image quality test objects embedded within the standard mammography phantom. Scoring consistency depended upon the difficulty of the individual test object scored, the scoring of partial test objects, and confusion in the scoring criteria. Test objects scored with artifact subtraction accounted for the largest source of variability. Scoring without artifact subtraction resulted in the most consistency, or least variability in phantom scores among the inspectors.