Abstract

Consensus scoring occurs when the scoring key for a test is based upon the responses of the norm group. Consensus scoring is an attractive alternative to traditional methods of creating a scoring key for ability tests, especially useful when experts disagree about the correct answers to test items, as they do in the area of emotions and emotion perception. Of the many variations of consensus scoring, mode consensus scoring (the most frequent response in a norm group is given a score of 1, and all other responses a score of 0) and proportion consensus scoring (each respondents score on an item is equal to the proportion of the norm group who match the respondents answer) are the most widely used and the most psychometrically promising. This paper demonstrates that mode consensus scoring is biased against smaller sub-groups within the norm group: when sub-groups differ in their modal responses, the size of the sub-groups will influence the average group score. No known scoring option eliminates this bias. In contrast, proportion consensus scoring is not necessarily biased against smaller groups, although bias does occur in some extreme situations. Proportion consensus scoring is therefore the preferred consensus scoring option at this time.