An instrument to measure medical students' perceptions of the assessment environment: Development and initial validation

Introduction: This study aimed to develop an instrument for measuring students' perceptions of the assessment environment in undergraduate medical programme and to examine the instrument's psychometric properties.

Method: The Assessment Environment Questionnaire (AEQ), a 40-item, 4-point Likert scale instrument (1=Strongly Disagree to 4=Strongly Agree) was developed and administered to medical undergraduates from the authors' institution from November 2013 to June 2014. The overall response rate was 626/794 (78.8%). To examine the construct validity of the instrument, the factor structure of the AEQ was determined through exploratory factor analysis with principal component analysis and varimax rotation. To check the internal consistency of the instrument and its factors/subscales, Cronbach's alpha coefficients were computed across all the items as well as items within each of the factor/subscale. Correlations within and between the different factors/subscales were computed to examine their convergent and divergent validity. Mean and standard deviation for each of the factor/subscale as well as the overall AEQ were computed. Possible influence(s) of selected demographic variables on the AEQ scores were examined.

Results: A total of 611 completed questionnaires were analysed. In initial analysis, six items with communalities below 0.40 were removed. Another item with rotated factor loading less than 0.50 was deleted. Eight items with cross factor loadings were also removed. A further five items in two subscales with alpha values of 0.44 and 0.52 were dropped. The 40-item AEQ was reduced to 20 items. In the final analysis, the Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis, with KMO=0.90, which is well above the acceptable limit of 0.50. The Bartlett's Test of sphericity $\chi^2(190)=4442.97$, $p<0.001$, indicated that correlations between items were sufficiently large for factor analysis to be conducted. Four factors with eigenvalues of 6.54, 1.93, 1.47 and 1.40 were retained, which together explained 56.72% of the variance. These four factors were labeled as: feedback mechanism (7 items), learning and performance (5 items), information on assessment (5 items) and assessment system/procedure (3 items). Reliability analysis reported an overall alpha value of 0.89, indicating the AEQ has good internal consistency. With alpha values ranging from 0.71 to 0.87, the four factors/subscales within the AEQ also have acceptable to high internal consistency, providing evidence for convergent validity of the items in each subscale. Item-total correlation was above 0.40 for 17 of the 20 items, with a mean of 0.50. This reflects good item discrimination, supporting the divergent validity of the instrument. Mean score for the AEQ was 2.68/4.00. The factor/subscale of `feedback mechanism' recorded the lowest mean (2.39/4.00) while the factor/subscale of `assessment system/procedure' scored the highest mean (2.92/4.00). Significant differences in means between respondents of various demographic backgrounds provided evidence to support the instrument's construct validity.

Conclusion: The AEQ appears to be a valid and reliable instrument. There was evidence of its construct validity, convergent and divergent validity. The instrument and its four subscales show good internal consistency. Initial validation of the instrument supports its use to measure students' perceptions of the assessment environment in an undergraduate medical programme.