Journal of Sports Sciences, 1999, **17**, 861–872

Development and validation of a mood measure for adolescents

PETER C. TERRY, ANDREW M. LANE,* HELEN J. LANE and LEE KEOHANE

Department of Sport Sciences, Brunel University, Osterley Campus, Borough Road, Isleworth, Middlesex TW7 5DU, UK

Accepted 3 May 1999

The aim of this study was to develop and validate a shortened version of the Profile of Mood States suitable for use with adolescents. The Profile of Mood States-Adolescents (POMS-A) was administered to 1693 participants from two populations: school children and young athletes. Confirmatory factor analysis supported the factorial validity of a 24-item six-factor model using both independent and multi-sample analyses. Correlations of POMS-A scores with previously validated inventories, which were consistent with theoretical predictions, provided evidence of criterion validity. It is proposed that the POMS-A is a valid instrument for the assessment of mood in adolescents.

Keywords: emotions, measurement, model testing, sport, structural equations.

Introduction

The quest to understand the psychology of emotion has generated persistent investigation of the construct of mood. Much attention has been paid to self-reported mood states and their attendant impact upon behaviour. Such research relies heavily upon parsimonious methods of assessing transient emotions. To date, psychometric development in this area has focused primarily upon adult populations, particularly students and psychiatric out-patients. The aim of this study was to develop and validate an inventory for assessing mood states that can be used with adolescent populations in the context of classroom and athletic environments. At least three arguments suggest the need for such a measure.

First, there has been considerable research to examine mood among adolescents. Mood has typically been assessed using the Profile of Mood States (POMS; McNair *et al.*, 1971). This profile describes six sub-components of the overall mood construct: anger, confusion, depression, fatigue, tension and vigour. The factor structure of the POMS and the associated tables of normative values were derived from groups of adults and psychiatric out-patients. In the test manual, the POMS is recommended for use with 'subjects aged 18 and older who have had at least some high school education' (McNair *et al.*, 1971, p. 6). No data from individuals under the age of 18 were used in the original validation studies and therefore the degree to which the POMS and its derivatives were suitable for research involving young participants is unknown.

Despite this limitation, the POMS has been used as a research tool with adolescents in physical education (Green *et al.*, 1995; Newcombe and Boyle, 1995), sport environments (Goss, 1994; Fry *et al.*, 1995) and clinical settings (Walker and Sprague, 1988). Furthermore, Hollander *et al.* (1995) proposed using the POMS to screen young athletes for mood changes as a precursor for overtraining syndrome (see Budgett, 1990). Given the research interest in mood in adolescents and the extent to which tests of theory rely upon valid measurement, demonstration of the construct validity of mood measures in the population of interest is an imperative.

Secondly, the original 65-item POMS has been criticized for taking too long to complete (Shacham, 1983; Grove and Prapavessis, 1992; Curren *et al.*, 1995), a criticism that also applies to the 72-item bipolar version (Lorr and McNair, 1984). This point is particularly relevant when mood is assessed in an ecologically valid setting, such as before competition or at the start of a lesson, where brevity is paramount. This has contributed to the development of several shortened versions of the profile (Shacham, 1983; Grove and

* Author to whom all correspondence should be addressed. e-mail: andrew.lane@brunel.ac.uk

Prapavessis, 1992; McNair *et al.*, 1992). However, it is important to recognize that completion time is influenced not only by the number of items but also their comprehensibility. Items in the original POMS and its shortened derivatives were generated from the responses of undergraduate students who were all aged 18 years or over. It has also been noted (Grove and Prapavessis, 1992) that some items have a distinct North American orientation, such as 'bushed' and 'blue', which may require explanation in other cultural contexts.

Thirdly, recent developments of computer software to test the factor structures of psychological questionnaires have prompted researchers to emphasize the benefits of structural modelling techniques such as confirmatory factor analysis (see Hendrick and Hendrick, 1986; Schutz and Gessaroli, 1993; Bentler, 1995; Thompson and Daniel, 1996). The traditional method used to demonstrate factorial validity has been exploratory factor analysis. It has previously been suggested that the replication of factors through exploratory analyses among disparate samples is evidence of factorial validity (Kerlinger, 1979; Gorsuch, 1983). However, the contemporary view is that a more rigorous procedure to test the generalizability of a measure is to use multi-sample confirmatory factor analysis to test the extent to which data support hypothesized relationships specified in a prior model across several different samples (Bentler, 1995; Tabachnick and Fidell, 1996; Thompson and Daniel, 1996). To date, neither the POMS (McNair *et al.*, 1971) nor its shortened versions (Shacham, 1983; Grove and Prapavessis, 1992; McNair *et al.*, 1992) have been scrutinized using either single-sample or multi-sample confirmatory factor analysis.

In summary, there is a need for a shortened version of the POMS developed specifically for younger populations. The aim of this study was to develop and validate such an inventory.

Research strategy

The validity of a self-report measure is defined as the 'degree to which a test or instrument measures what it purports to measure' (Thomas and Nelson, 1990, p. 527). A self-report measure is considered valid when it has demonstrated content validity, factorial validity, criterion validity and construct validity (see American Psychological Association, 1974; Thomas and Nelson, 1990; Anastasi and Urbina, 1997). According to Anastasi and Urbina (1997, p. 114), construct validity is of paramount importance, and can be seen as 'inclusive validity, insofar as it specifies what the test measures'. Content, factorial and criterion validation procedures are among the sources of information that contribute to the definition and understanding of the constructs assessed by the self-report measure and, therefore, act as the basis by which construct validity is judged.

An important research decision in the development of a questionnaire is the number of items included in each factor, particularly when brevity is important. It is suggested that, theoretically, an infinite number of items is potentially available for the measurement of any construct (Anastasi and Urbina, 1997). Consequently, the strategy for identifying a construct typically starts with a relatively large pool of items that is reduced through subsequent analyses. Kline (1993) cautioned that item reduction might yield a factor containing items with an extremely similar meaning. He suggested that this may lead to two problems with the validity of the factor. The first is that the factor might show high validity coefficients and thereby show evidence of construct validity. Secondly, the factor might assess a limited dimension of the construct; therefore, the factor should be re-labelled to reflect the dimension of the construct it is assessing. In the present study, therefore, validation of the inventory was done over a series of different stages with each stage acting as a check on the findings from the previous stage.

As the aim of the study was to develop a short questionnaire, this raises the question of the optimum number of items needed to assess a construct. Jackson and Marsh (1996) argued that the optimum number of items needed to describe a construct in a short questionnaire is four; Bollen (1989) cautioned against reducing the number of items in a factor to less than three. From a statistical perspective, Watson and Clark (1997) reported that factors with less than four items typically fail to yield an internal consistency (alpha) coefficient (Cronbach, 1951) above the generally accepted criterion value of 0.70 (Tabachnick and Fidell, 1996). As alpha coefficients are influenced by the number of items in a factor, it is suggested that each factor of a questionnaire should contain the same number of items to facilitate accurate comparisons of internal consistency. We therefore wished to produce a version of the POMS with six factors of four items each.

The research process had three stages. First, to establish content validity, the suitability of an initial item pool was assessed by a panel of experts and by school children. Secondly, to establish factorial validity, the hypothesized factor structure of the item pool was assessed using confirmatory factor analysis of the mood responses of school children. A revised model was then tested among samples of school children and young athletes simultaneously using multi-sample confirmatory factor analysis. The third stage, to establish criterion validity, tested the extent to which the subscales of the questionnaire correlated with previously validated measures.

Stage 1: Content validity

Content validity refers to the extent to which items represent the construct they are purported to measure. A standard approach to establishing content validity is to use experts (McNair *et al.*, 1971; Martens *et al.*, 1990), a representative sample of participants (Jones *et al.*, 1990) or both (Widmeyer *et al.*, 1985) to select or confirm items that best describe the construct in question. The preliminary stages in developing the Profile of Mood States-Adolescents (POMS-A) involved experts and school children in the process of refining the selection of appropriate mood descriptors.

An initial item pool of 83 mood descriptors was established consisting of the 65 adjectives derived from the original POMS plus, where it was suspected that existing items might prove inappropriate for adolescents, 18 additional adjectives taken from a thesaurus. Ten teachers of English at secondary schools identified those adjectives whose meaning they believed 'would be understood by the vast majority of children in the 14–16 age group'. Items were eliminated if four or more teachers identified them as inappropriate, resulting in 13 items being discarded.

To maximize comprehensibility among children, and retaining the original conceptualization of the mood construct, a sample of 50 children (age 14–15 years) rated the extent to which the remaining 70 items described the original factors (anger, confusion, depression, fatigue, tension and vigour). Participants identified items that, according to their understanding, were closest in meaning to the six mood factors. The top seven items under each of the six headings were selected for an initial 42-item inventory and were assigned randomly to order. The rationale for using the top seven items was to retain a sufficient number of items to allow further item reduction after confirmatory factor analysis among a larger sample. We also judged, based on the characteristics of the Cronbach alpha estimates described earlier, that each factor should contain the same number of items.

Stage 2.1: Preliminary test of factorial validity

Confirmatory factor analysis was used to test factorial validity. Schutz (1994) argued that, when researchers have a hypothesized model to test, the first test of factorial validity should be confirmatory. As confirmatory factor analysis is a test of theory, it is important to assess also the theoretical integrity of the proposed item groupings. To this end, the nature of the six factors is described and proposed relationships among factors are hypothesized.

Anger is typified by feelings that vary in intensity from mild annoyance or aggravation to fury and rage, and is associated with arousal of the autonomic nervous system (Spielberger, 1991). Confusion is proposed to be a feeling state characterized by feelings of bewilderment and uncertainty, and is associated with a general failure to control attention and emotions. Depression is associated with a negative self-schema characterized by themes such as hopelessness, personal deficiency, worthlessness and self-blame (Beck and Clark, 1988). Fatigue is typified by feelings of mental and physical tiredness. Tension is typified by such feelings as nervousness, apprehension, worry and anxiety. Finally, vigour is typified by feelings of excitement, alertness and physical energy.

We hypothesized that depression would show moderate positive relationships with anger, confusion, fatigue and tension, and a weak inverse relationship with vigour. We also expected vigour to show a moderate inverse relationship with fatigue and be unrelated to anger, confusion and tension. This pattern of intercorrelations among mood dimensions has been found with both athletes (Grove and Prapavessis, 1992; Terry and Slade, 1995) and students (McNair *et al.*, 1971) when mood is assessed using the 'How do you feel right now' response set. A general limitation of mood research in sport is that intercorrelations among POMS dimensions have rarely been reported.

The research strategy at this stage was to use confirmatory factor analysis to refine the 42-item six-factor model by removing the three psychometrically weakest items in each scale, and thereby to produce an instrument of greater psychometric integrity. Model testing involved two stages: the first stage tested the 42-item six-factor model, and the second stage tested a revised 24-item six-factor model. It could be argued that the research strategy should test alternative models of mood, rather than refine the six-factor model. In this study, alternative models of mood would be explored at a later stage if the POMS model of mood was rejected.

Methods

Participants. Students at a secondary school in the suburbs of west London, England, participated in a study to test the factor structure of an initial 42-item and a revised 24-item version of the Profile of Mood States-Adolescents (POMS-A). The school population had a mixed socio-economic and ethnic composition. Most secondary schools in England teach students in the age range 11–18 years.

The sample comprised 416 children aged 14–16 years (mean $\pm$ *s*: 14.8 $\pm$ 1.0 years). Forty-eight participants failed to respond to one or more items and their responses were discarded. Therefore, 368 completed

questionnaires from 199 males and 169 females were available for analysis.

Procedure. The inventory was completed by participants in a classroom setting. They were asked to rate 'How are you feeling right now' in terms of mood descriptors such as 'worried' and 'unhappy'. Responses were provided on a scale from 0 ('not at all') to 4 ('extremely'). To ensure consistency during data collection, instructions were read from a prepared script. Furthermore, a culturally appropriate alternative word list (cf. Albrecht and Ewing, 1989) was available to participants for reference in case mood descriptors could not be understood, although no participant referred to this list.

Data analysis

Confirmatory factor analysis using EQS V5 (Bentler and Wu, 1995) was used to test the 42-item six-factor structure of mood. As multivariate non-normality was evidenced (Mardia's value = 117.5), the data were analysed using the maximum likelihood robust estimation method (see Hoyle, 1995). The model specified that items were related to their hypothesized factor with the variance of the factor fixed to 1. Anger, confusion, depression, fatigue and tension were allowed to intercorrelate. Vigour was allowed to correlate with depression and fatigue only, as it was hypothesized that the vigour–anger, vigour–confusion and vigour–tension relationships would not differ significantly from zero.

According to Hu and Bentler (1995), there is little agreement among researchers about the best index of the overall fit of a model tested by confirmatory factor analysis. Consequently, it has been suggested that researchers should report several different fit indices (Hoyle and Panter, 1995). First, the chi-squared to degrees-of-freedom ratio ($\chi^2 : df$) ratio has been proposed as a superior index to χ^2 because, with large samples and complex models, there is a tendency for χ^2 values to be inflated, causing good fitting models to be rejected erroneously (Byrne, 1989). Although Byrne (1989) suggested that a $\chi^2 : df$ ratio of <2 indicates an acceptable fit, researchers have suggested examining fit indices which use the χ^2 in conjunction with other fit indices (see Hu and Bentler, 1995, for a review).

Other fit indices used to assess the model included the goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI). These two indices 'assess the relative amount of the observed variance and covariances accounted for by the model' and, as such, are analogous to the R^2 typically used in multiple regression (Hoyle and Panter, 1995). The adjusted goodness-of-fit index is similar to the goodness-of-fit index but takes into account the complexity of the model. Hu and Bentler (1995) contend that the goodness-of-fit index is the most reliable absolute fit index. The criterion value for an acceptable fit is 0.90 for both indices.

Two incremental fit indices were also used to judge factorial validity, the non-normed fit index (NNFI: Tucker and Lewis, 1973) and the robust comparative fit index (RCFI: Bentler, 1990). The non-normed fit index takes into account sample size and is thus suggested to provide a better estimate of the fit of a model than the normed fit index when multivariate normality is violated. The robust comparative fit index evaluates the adequacy of the hypothesized model in relation to the worst (independent) model. If the hypothesized model is not a significant improvement on the independent model, the fit index will be close to zero (Bentler, 1995). These indices have been found to control effectively for overestimation of χ^2, underestimation of incremental fit indices, and under-identification of errors when data are not normally distributed (see West *et al.*, 1995). The criterion value for both incremental fit indices is 0.90 (Bentler, 1995). Finally, the root mean square error of approximation (RMSEA; Steiger, 1990) was used as a measure of the extent to which the model was supported per degree of freedom. Browne and Cudeck (1993) proposed that a value of 0.05 or lower indicates a close fit, and values up to 0.08 represent a reasonable fit.

Multiple selection criteria were used in the process of identifying the best four items for each scale: (1) the highest four-factor loadings, (2) the lowest error measurements and (3) a Cronbach alpha coefficient above 0.70 for the resultant factor. Previous research has used a similar strategy to reduce the number of items to produce a more parsimonious version of an existing questionnaire (see Jackson and Marsh, 1996).

Results and discussion

The aim of this stage of the research was to test a 42-item and a 24-item six-factor model of mood using confirmatory factor analysis. Results for the 42-item model showed poor fit (goodness-of-fit index = 0.780, adjusted goodness-of-fit index = 0.754, non-normed fit index = 0.823, robust comparative fit index = 0.854) except the Satorra-Bentler scaled $\chi^2 : df$ ratio of 1.48. The standardized solution results (see Table 1) offered support for the hypothesized relationships between most of the items, although several items showed a weak relationship with their hypothesized factor ('forgetful', 'lonely', 'ready to fight', 'spiteful' and 'stimulated').

Following removal of the weakest three items from each scale, the results of confirmatory factor analysis of the revised 24-item POMS-A yielded acceptable fit indices for the non-normed fit index (0.92), robust comparative fit index (0.93) and root mean square error of approximation (0.06). The goodness-of-fit (0.89) and

Table 1 Standardized solution for factor loadings of the 42-item version of the POMS-A

Item	Factor loading	Error variance
Worried	0.738	0.675
Nervous	0.716	0.698
Anxious	0.655	0.755
Panicky	0.572	0.820
On edge	0.558	0.830
Tense	0.528	0.849
Stressed	0.524	0.852
Unhappy	0.831	0.556
Miserable	0.807	0.590
Depressed	0.763	0.646
Downhearted	0.684	0.730
Sad	0.648	0.762
Gloomy	0.624	0.781
Lonely	0.398	0.918
Energetic	0.872	0.489
Active	0.868	0.496
Lively	0.725	0.689
Alert	0.607	0.795
Cheerful	0.468	0.884
Brisk	0.417	0.909
Stimulated	0.370	0.929
Mixed-up	0.847	0.532
Confused	0.708	0.707
Uncertain	0.707	0.707
Muddled	0.677	0.736
Uneasy	0.513	0.858
Unable to concentrate	0.372	0.928
Forgetful	0.250	0.968
Angry	0.805	0.593
Bitter	0.756	0.654
Annoyed	0.755	0.656
Bad-tempered	0.692	0.722
Furious	0.651	0.759
Spiteful	0.432	0.902
Ready to fight	0.415	0.910
Tired	0.887	0.461
Sleepy	0.848	0.529
Exhausted	0.786	0.618
Worn-out	0.782	0.624
Ready for bed	0.710	0.704
Bushed	0.694	0.720
Weary	0.539	0.842

adjusted goodness-of-fit (0.86) indices were marginally below the 0.90 criterion value. Cronbach alpha coefficients all exceeded the 0.70 criterion value (see Table 2). Collectively, the results indicated that the 24-item six-factor model should go forward to the next stage of validation.

Stage 2.2: Test of the generalizability of the factor structure

The establishment of factorial validity involves demonstrating that the hypothesized factor structure can be replicated in disparative samples. Only when such consistency has been demonstrated can findings justifiably be extrapolated to other populations (Taylor, 1987). In the present study, validity testing was extended to a new sample of school children and a sample of young athletes. Confirmatory factor analysis was used to test the extent to which the data from the two samples supported the relationships specified in the 24-item six-factor model.

Methods

Participants. Sample 1 comprised 683 young athletes (301 males, 382 females; mean $\pm$ *s*: age 14.7 $\pm$ 1.8 years), who participated in the sports of archery, field hockey, judo, netball, soccer, table-tennis, track and field, trampolining, triathlon and volleyball. Sample 2 comprised 594 school children aged 14.7 $\pm$ 1.4 years (339 males, 313 females).

Procedure. The POMS-A was administered 1 h before competition for the athletic sample and at the start or the end of a class lesson for the school children. The remaining procedures used replicated those reported above.

Results and discussion

The χ^2:*df* ratio indicated an acceptable fit of the data to the model in both samples: 2.06 for school children and 2.19 for young athletes. The fit indices provided further support for the fit of the model in both samples, with fit indices above 0.90 for the goodness-of-fit index (school children = 0.908, young athletes = 0.905), non-normed fit index (school children = 0.919, young athletes = 0.901) and robust comparative fit index (school children = 0.925, young athletes = 0.912); adjusted goodness-of-fit indices were marginally below the 0.90 criterion value (school children = 0.884, young athletes = 0.881). Root mean square errors of approximation were also acceptable (school children = 0.052, young athletes = 0.062). Alpha coefficients ranged from 0.75 to 0.86, indicating that the factors contain items which are internally consistent (see Table 2).

The standardized solution showed that all factor loadings were higher than 0.50 except 'anxious' (0.456) in the sample of young athletes and 'alert' (0.491) in the sample of school children. Correlation coefficients among mood dimensions were consistent with the

Table 2 Internal consistency estimates of mood dimensions among three samples

	Sample A: school children (n = 369)			
	42 items	24 items	Young athletes (n = 683)	School children (n = 594)
Anger	0.80	0.82	0.80	0.80
Confusion	0.76	0.83	0.86	0.81
Depression	0.86	0.85	0.85	0.85
Fatigue	0.90	0.90	0.82	0.85
Tension	0.82	0.74	0.75	0.82
Vigour	0.82	0.85	0.79	0.79

Table 3 Inter-correlations of POMS-A subscales among school children (n = 594) and young athletes (n = 683)

	Anger	Confusion	Depression	Fatigue
Confusion				
Young athletes	0.588			
School children	0.613			
Depression				
Young athletes	0.876	0.700		
School children	0.802	0.731		
Fatigue				
Young athletes	0.363	0.388	0.285	
School children	0.275	0.309	0.382	
Tension				
Young athletes	0.186	0.492	0.292	0.174
School children	0.316	0.511	0.443	0.250
Vigour				
Young athletes			−0.079	−0.288
School children			−0.145	−0.357

hypothesized model (see Table 3). Depression showed moderate-to-strong positive relationships with anger, confusion, fatigue and tension, and a weak inverse relationship with vigour. Lagrange Multiplier Test results indicated that the fit of the model would not be improved by allowing vigour to correlate with anger, confusion and tension.

The strength of the intercorrelations among factors was weaker than those typically reported by researchers using the original POMS (see McNair *et al.*, 1971, 1992; Grove and Prapavessis, 1992; Terry and Slade, 1995). This suggests that the POMS-A shows greater factorial independence than the original version. Collectively, the results provide strong support for the factorial validity of the 24-item POMS-A in the two samples independently. The next step in the validation process was to test the hypothesized model using multi-sample analysis.

Stage 2.3: Multi-sample analysis

Multi-sample confirmatory factor analysis was used to determine the strength of the factor solution across the samples of young athletes (n = 683) and school children (n = 594) simultaneously. In multi-sample analysis, it is assumed that data from more than one sample provide comparable information about the hypothesized model. This assumption is tested by analysing data from different samples simultaneously to verify the extent to which the model reproduces the data of each sample to within sampling accuracy (see Bentler, 1995). As with one-sample confirmatory factor analysis, χ^2 statistics, the goodness-of-fit index, adjusted goodness-of-fit index, non-normed fit index and comparative fit index represent the extent to which variance–covariance matrices from different samples are identical. It is

Table 4 POMS-A scores among school children and young athletes (mean ± *s*)

	School children (n = 594)	Young athletes (n = 683)	F_{1271}	Effect size
Anger	1.52 ± 2.53	1.26 ± 2.30	3.47	0.11
Confusion	2.12 ± 2.85	1.66 ± 2.54	9.68*	0.17
Depression	1.90 ± 2.98	1.36 ± 2.58	8.21*	0.20
Fatigue	5.07 ± 3.77	3.23 ± 3.19	41.6*	0.53
Tension	2.85 ± 3.46	3.63 ± 3.17	17.6*	0.24
Vigour	6.90 ± 4.05	8.76 ± 4.14	65.4*	0.45
	Wilks' $\lambda_{6,1266} = 0.91$, $P < 0.001$			

* $P < 0.001$.

important to note that ESQ V5 does not give robust estimates in multi-sample analysis.

In multi-sample confirmatory factor analysis, it is possible to test several different hypotheses regarding the similarity of relationships across samples using the Lagrange Multiplier Test. In multi-sample analysis, this test examines the extent to which the fit of the model would be improved if equality constraints were removed. Following the recommendations of Bentler (1995), a hierarchical procedure was used to place equality constraints on hypothesized relationships. The first multi-sample analysis tested the model with no equality to get a baseline score on which to compare more restricted models. The second analysis placed equality constraints on factor loadings. The third analysis placed equality constraints on factor loadings and correlation coefficients between factors. We hypothesized that equality constraints on all relationships would hold between the two groups.

Before conducting multi-sample analyses, differences in the intensity of mood responses between the two samples were examined. Demonstration of differences in the intensity of mood lends support to the notion that the participant groups derive from two different populations. Factor scores were calculated by summing item scores within each factor; multivariate analysis of variance (MANOVA) was then used to assess differences in the intensity of factor scores between the two samples. This MANOVA showed significant differences in the intensity of mood responses between school children and young athletes (Wilks' $\lambda_{6,1266} = 0.91$, $P < 0.001$; see Table 4). Univariate differences indicated that the school children reported significantly more confusion, depression and fatigue, but less tension and vigour, than the young athletes. Calculation of effect sizes showed that the group differences were relatively small for confusion, depression and tension, but moderate for fatigue and vigour (see Table 4).

Differences between school children and young athletes in the intensity of reported mood were consistent with results from a previous large-scale between-group comparison of mood responses (Terry and Lane, in press), which found that adult athletes reported lower anger, confusion, depression, fatigue and tension than adult students but higher vigour scores. The higher tension scores reported by the young athletes in the present study can be attributed to the assessment of mood before competition. Terry and Lane (in press) also found that athletes reported higher tension scores before competition than when away from the competition environment. Abele and Brehm (1993) reported that mood changes in competitive sport are typified by a decrease in tension scores from the beginning to the end of competition. Collectively, the results of the MANOVA confirmed significant variations in reported mood between school children and young athletes, and thus support the notion that the two groups represented disparate populations even though most of the young athletes would also have been school children.

Results and discussion

The results of multi-sample confirmatory factor analysis are given in Table 5. Results supported the baseline model (goodness-of-fit index = 0.906, comparative fit index = 0.922) and the model which constrained factor loadings to be equal (goodness-of-fit index = 0.901, comparative fit index = 0.918). A test of the extent to which relationships among factors, and relationships between items and factors, were equal in the two groups showed acceptable fit indices (goodness-of-fit index = 0.900, comparative fit index = 0.916).

The Lagrange Multiplier Test showed that 3 of 36 equality constraints differed between the two samples. Specifically, relationships between depression and fatigue ($\chi^2 = 12.21$, $P < 0.01$) and factor loadings for

Table 5 Multi-sample confirmatory factor analysis of the POMS-A among school children and young athletes

Fit statistics	Unconstrained model ($df = 480$)	Factor loadings ($df = 540$)	Factor loadings and correlations ($df = 516$)
$\chi^2 : df$ ratio	3.275	3.055	3.254
Goodness-of-fit index	0.906	0.901	0.900
Adjusted goodness-of-fit index	0.941	0.941	0.942
Non-normed fit index	0.910	0.910	0.911
Comparative fit index	0.922	0.918	0.916
Root mean square error of approximation	0.053	0.042	0.067

'annoyed' ($\chi^2 = 18.32$, $P < 0.01$) and 'worn-out' ($\chi^2 = 9.83$, $P < 0.01$) differed significantly. Although these statistics may appear to question the generalizability of the solution, further analysis with the equality constraints removed indicated only a marginally improved fit of the model (see Table 5).

A multi-sample confirmatory factor analysis with constraints imposed on the hypothesized relationships is proposed to provide a very rigorous test of factorial validity (Bentler, 1995). The results of the present study show that hypothesized relationships remain stable between different groups of participants even when the intensity of mood responses differ significantly. Collectively, it is proposed that the POMS-A shows strong evidence of factorial validity for use with school children and young athletes.

Stage 3: Test of criterion validity

Criterion validity is defined as the 'degree to which scores on a test are related to some recognised standard, or criterion' (Thomas and Nelson, 1990, p. 516). Criterion validity can be demonstrated using either concurrent validity or predictive validity. Concurrent validity is defined as when 'a measuring instrument is correlated with some criterion that is administered at the same time, or concurrently' (Thomas and Nelson, 1990, p. 515). Concurrent validity is typically examined by correlating scores from two questionnaires that were administered concurrently, with the previously validated questionnaire being the criterion measure (see McNair *et al.*, 1971, 1992). Concurrent validity is inferred from the strength of correlations between two questionnaires which measure similar constructs.

The criterion measure for the vigour scale was the 'positive affect' scale from the Positive and Negative Affect Schedule (Watson *et al.*, 1988). The criterion measure for the anger, confusion, depression, fatigue and tension scales was the 'negative affect' scale from the Positive and Negative Affect Schedule. A further criterion measure for anger was provided by the State–Trait Anger Expression Inventory (Spielberger, 1991).

Methods

Participants. A total of 182 participants completed the POMS-A and a second questionnaire. Ninety-one participants (age 14.3 ± 1.2 years) completed the Positive and Negative Affect Schedule and the POMS-A. The other 91 participants (age 13.3 ± 0.9 years) completed the State–Trait Anger Expression Inventory and the POMS-A. Participants were school children at a mixed secondary school in North West London.

Measures of criterion validity. Watson *et al.* (1988) developed the Positive and Negative Affect Schedule to assess independent markers of positive and negative affect. Items are rated on a 5-point scale anchored by 'not at all' (1) and 'extremely' (5). Examples of positive affect items include 'excited', 'enthusiastic' and 'determined'; examples of negative affect items include 'distressed', 'guilty' and 'scared'. The validation studies for the Positive and Negative Affect Schedule, which involved 3554 completions of the inventory, demonstrated strong content validity with all items loading at 0.50 or higher onto their hypothesized factor. Cronbach alpha coefficients ranged from 0.84 to 0.90. Recent research has confirmed the factor structure of this schedule among young athletes (Crocker, 1997). With reference to the present study, this suggests that this schedule is both valid for use in sport and valid for use with children. Consequently, it is an excellent research tool to test the criterion validity of the POMS-A.

The factor structure of the 10-item State–Trait Anger Expression Inventory was validated by Spielberger (1991) using a sample of 550 individuals. Factor analysis yielded a single factor with a Cronbach alpha coefficient of 0.93. Items are rated on a 4-point scale anchored by 'almost never' (1) and 'very often' (4).

Table 6 Correlations between POMS-A scores and criterion measures

Criterion measure	Anger	Confusion	Depression	Fatigue	Tension	Vigour
State anger						
Expression inventory	0.82*	0.23	0.35*	0.36*	0.29*	−0.03
Positive affect	0.16	0.13	0.10	0.01	0.21	0.62*
Negative affect	0.78*	0.80*	0.75*	0.52*	0.72*	−0.01

* $P < 0.01$.

Procedure. The participants completed the questionnaires in accordance with the procedure used in Stages 2.1 and 2.2. Pearson product–moment correlation was used to assess the relationship between POMS-A scores and the criterion measure.

Results and discussion

Correlation coefficients between the POMS-A measures and the criterion measures are shown in Table 6. Consistent with the hypothesized predictions, scores on the POMS-A anger scale showed the strongest correlation with State–Trait Anger Expression Inventory scores, and vigour correlated significantly with positive affect. Anger, confusion, depression, fatigue and tension correlated significantly with negative affect. Also consistent with our theoretical proposals, negative affect did not correlate significantly with vigour, and positive affect did not correlate significantly with the other POMS-A factors. Consistent with previous research, negative affect showed a moderate relationship with anger, confusion, depression, fatigue and tension (Watson *et al.*, 1988). Collectively, the results suggest that the POMS-A shows evidence of concurrent validity.

General discussion

The aim of this study was to develop and validate a shortened version of the POMS suitable for use with children in the context of classroom and athletic environments. Theory testing and construct measurement are inextricably linked. It has been suggested that the development of a valid measure of the theoretical construct to be examined should be the first step in the research process (Hendrick and Hendrick, 1986). If the construct validity of the instrument is questionable, then it is not possible to accurately test the theory under investigation. Watson *et al.* (1988) argued that poorly developed and validated mood inventories contributed to the ambiguity of research findings on the nature of mood in the psychology literature. The three-stage validation process used here tested content validity, factorial validity and criterion validity and has provided strong support for the 24-item version of the POMS. Collectively, it is proposed that the POMS-A shows clear evidence of construct validity. However, as validation is an ongoing process (Anastasi and Urbina, 1997), we suggest that future research should continue to investigate the validity of the POMS-A. A limitation of the present concurrent validity process was that the State–Trait Anger Expression Inventory criterion measures were validated on samples from student rather than athletic populations. Therefore, we suggest that researchers investigate the validity of such measures as the State–Trait Anger Expression Inventory (Spielberger, 1991) in sport. Indeed, there is a need for more thorough validation in the sport environment of many of the inventories used in sport psychology research, which have been validated on other populations (Schutz, 1994).

Comparison with the original POMS

Of the 24 items on the POMS-A, 17 are also on the original POMS: 'active', 'alert', 'angry', 'annoyed', 'anxious', 'bad-tempered', 'bitter', 'confused', 'energetic', 'exhausted', 'lively', 'miserable', 'muddled', 'nervous', 'panicky', 'uncertain', 'unhappy' and 'worn-out'. Fourteen items on the original POMS were discarded from the POMS-A: 'bushed', 'cheerful', 'forgetful', 'furious', 'gloomy', 'lonely', 'on-edge', 'ready to fight', 'sad', 'spiteful', 'tense', 'unable to concentrate', 'uneasy' and 'weary'. Previous research has questioned whether athletes interpret items such as 'ready to fight' literally (Terry and Slade, 1995). Additionally, 'cheerful' appears to be part of a construct labelled happiness, rather than vigour. In the present study, vigour comprises items associated with positive feelings and high arousal (active, alert, energetic and lively), which appear to be more appropriate descriptors of the vigour construct.

Uses of the POMS-A

Recent research has seen the development of a conceptual model to explain relationships between mood

and athletic performance (Lane and Terry, 1998). The conceptual model proposes that mood should be measured through the six mood dimensions identified in the POMS. Lane and Terry (1998) suggested that depressed mood was associated with increased anger, confusion, fatigue and tension, and reduced vigour. In addition, depressed mood was also proposed to moderate mood and performance relationships for anger and tension. We suggest, therefore, that the POMS-A should be used to investigate the hypotheses made in the conceptual model proposed by Lane and Terry (1998).

The POMS-A may provide a useful tool for applied sport psychology research, as it is short, easy to complete and suitable for use with adolescents. Consequently, the POMS-A can be used to assess mood shortly before competition without disturbing to a great extent athletes' normal pre-competition routines. The brevity of the questionnaire also means that it provides an instrument for assessing mood changes in education settings and for screening young athletes for mood disturbance as a precursor to overtraining syndrome. The POMS-A is appended to this paper; researchers are invited to use the scale without written permission from the authors.

Conclusions

The aim of this study was to develop and validate a measure of mood for adolescents. Confirmatory factor analysis supported the factorial validity of a 24-item six-factor model using both independent and multi-sample analyses. Criterion validity was demonstrated by correlating POMS-A scores with previously validated inventories. It is proposed that the POMS-A demonstrates construct validity for the assessment of mood in 11- to 18-year-olds. We suggest that future research should extend the validation process of the POMS-A to adults and adult athletes.

References

Abele, A. and Brehm, W. (1993). Mood effects of exercise versus sports games: Findings and implications for well-being and health. *International Journal of Health Psychology*, **2**, 53–81.

Albrecht, R.R. and Ewing, S.J. (1989). Standardizing the administration of the Profile of Mood States (POMS): Development of alternative word lists. *Journal of Personality Assessments*, **53**, 31–39.

American Psychological Association (1974). *Standards for Educational Psychological Testing*. Washington, DC: APA.

Anastasi, A. and Urbina, S. (1997). *Psychological Testing*, 7th edn. New York: Prentice-Hall.

Beck, A.T. and Clark, D.A. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research*, **1**, 23–56.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, **107**, 238–246.

Bentler, P.M. (1995). *EQS Structural Equation Program Manual*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P.M. and Wu, E.J.C. (1995). *EQS/Windows User's Guide*. Los Angeles, CA: BMDP Statistical Software.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Browne, M.W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In *Testing Structural Equation Models* (edited by K.A. Bollen and J.S. Long), pp. 132–162. Newbury, CA: Sage.

Budgett, R. (1990). Overtraining syndrome. *British Journal of Sports Medicine*, **24**, 231–236.

Byrne, B.M. (1989). *A Primer of Lisrel: Basic Applications Programming for Confirmatory Factor Analytic Models*. New York: Springer-Verlag.

Crocker, P.R.E. (1997). A confirmatory factor analysis of the Positive Affect Negative Affect Schedule (PANAS) with a youth sport sample. *Journal of Sport and Exercise Psychology*, **19**, 91–97.

Cronbach, L.J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika*, **16**, 297–334.

Curren, S.L., Andrykowski, M.A. and Studts, J.L. (1995). Short form of the Profile of Mood States (POMS-SF): Psychometric information. *Psychological Assessment*, **7**, 80–83.

Fry, A.C., Stone, M.H., Thrush, J.T. and Fleck, S.J. (1995). Precompetition training sessions enhance competitive performance of high anxiety junior weightlifters. *Journal of Strength and Conditioning Research*, **9**, 37–42.

Gorsuch, R.L. (1983). *Factor Analysis*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum.

Goss, J.D. (1994). Hardiness and mood disturbance in swimmers while overtraining. *Journal of Sport and Exercise Psychology*, **16**, 135–149.

Green, E.K., Burke, K.L., Nix, C.L., Lambrecht, K.W. and Mason, C. (1995). Psychological factors associated with alcohol use by high school athletes. *Journal of Sport Behavior*, **18**, 195–208.

Grove, J.R. and Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated Profile of Mood States. *International Journal of Sport Psychology*, **23**, 93–109.

Hendrick, C. and Hendrick, S. (1986). A theory and method of love. *Journal of Personality and Social Psychology*, **50**, 392–402.

Hollander, D.B., Meyers, M.C. and LeUnes, A. (1995). Psychological factors associated with overtraining: Implications for youth sport coaches. *Journal of Sport Behavior*, **18**, 3–20.

Hoyle, R.H. (ed.) (1995). *Structural Equation Modeling: Concepts, Issues, and Applications*. London: Sage.

Hoyle, R.H. and Panter, A.T. (1995). Writing about structural equation models. In *Structural Equation Modeling: Concepts, Issues, and Applications* (edited by R.H. Hoyle), pp. 158–175. London: Sage.

Hu, L. and Bentler, P.M. (1995). Evaluating model fit. In *Structural Equation Modeling: Concepts, Issues, and Applications* (edited by R.H. Hoyle), pp. 76–99. London: Sage.

Jackson, S.A. and Marsh, H.W. (1996). Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of Sport and Exercise Psychology*, **18**, 17–35.

Jones, J.G., Swain, A.B.J. and Cale, A. (1990). Antecedents of multidimensional state anxiety and self-confidence in elite intercollegiate middle-distance runners. *The Sport Psychologist*, **4**, 107–118.

Kerlinger, F.N. (1979). *Behavioral Research: A Conceptual Approach*. New York: Holt, Rinehart & Winston.

Kline, P. (1993). *Handbook of Psychological Testing*. London: Routledge.

Lane, A.M. and Terry, P.C. (1998). Mood states as predictors of performance: A conceptual model. *Journal of Sports Sciences*, **16**, 87.

Lorr, M. and McNair, D. (1984). *Manual for the Profile of Mood States Bipolar Form*. San Diego, CA: Educational and Industrial Testing Service.

Martens, R., Vealey, R.S., Burton, D., Bump, L. and Smith, D.E. (1990). Development and validation of the Competitive Sports Anxiety Inventory 2. In *Competitive Anxiety in Sport* (edited by R. Martens, R.S. Vealey and D. Burton), pp. 117–178. Champaign, IL: Human Kinetics.

McNair, D.M., Lorr, M. and Droppleman, L.F. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

McNair, D.M., Lorr, M. and Droppleman, L.F. (1992). *Revised Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Newcombe, P.A. and Boyle, G.J. (1995). High school students' sports personalities: Variations across participation level, gender, type of sport, and success. *International Journal of Sport Psychology*, **26**, 277–294.

Schutz, R.W. (1994). Methodological issues and measurement problems in sport psychology. In *International Perspectives on Sport and Exercise Psychology* (edited by S. Serpa, J. Alves and V. Pataco), pp. 35–57. Morgantown, WV: Fitness Information Technology.

Schutz, R.W. and Gessaroli, M.E. (1993). Use, misuse, and disuse of statistics in psychology research. In *Handbook of Research on Sport Psychology* (edited by R.N. Singer, M. Murphey and L.K. Tennant), pp. 901–921. New York: Macmillan.

Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of Personality Assessment*, **47**, 305–306.

Spielberger, C.D. (1991). *Manual for the State–Trait Anger Expression Inventory*. Odessa, FL: Psychological Assessment Resources.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioural Research*, **25**, 173–180.

Tabachnick, B.G. and Fidell, L.S. (1996). *Using Multivariate Statistics*. New York: Harper & Row.

Taylor, J. (1987). A review of validity issues in sport psychological research: Types, problems, solutions. *Journal of Sports Behavior*, **10**, 3–15.

Terry, P.C. and Lane, A.M. (in press). Development of normative data for the Profile of Mood States for use with athletic samples. *Journal of Applied Sport Psychology*.

Terry, P.C. and Slade, A. (1995). Discriminant capability of psychological state measures in predicting performance outcome in karate competition. *Perceptual and Motor Skills*, **81**, 275–286.

Thomas, J.R. and Nelson, J.K. (1990). *Research Methods in Physical Activity*. Champaign, IL: Human Kinetics.

Thompson, B. and Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, **56**, 197–208.

Tucker, L.R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, **38**, 1–10.

Walker, M.K. and Sprague, R.L. (1988). Effects of methylphenidate hydrochloride on the subjective reporting of mood in children with attention deficit disorder. Special Issue: Interface between mental and physical illness. *Issues in Mental Health Nursing*, **9**, 373–385.

Watson, D. and Clark, L.A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment*, **68**, 267–296.

Watson, D., Clark, L.A. and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, **54**, 1063–1070.

West, S.G., Finch, J.F. and Curren, P.J. (1995). Structural equations models with nonnormal variables: Problems and remedies. In *Structural Equation Modelling: Concepts, Issues, and Applications* (edited by R.H. Hoyle), pp. 56–73. London: Sage.

Widmeyer, W.N., Brawley, L.R. and Carron, A.V. (1985). The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport Psychology*, **7**, 244–266.

Appendix: The Profile of Mood States-A

Below is a list of words that describe feelings that people have. Please read each one carefully. Then circle the answer which best describes HOW YOU FEEL RIGHT NOW. Make sure you answer every question.

	Not at all	A little	Moderately	Quite a bit	Extremely
1. Panicky	0	1	2	3	4
2. Lively	0	1	2	3	4
3. Confused	0	1	2	3	4
4. Worn out	0	1	2	3	4
5. Depressed	0	1	2	3	4
6. Downhearted	0	1	2	3	4
7. Annoyed	0	1	2	3	4
8. Exhausted	0	1	2	3	4
9. Mixed-up	0	1	2	3	4
10. Sleepy	0	1	2	3	4
11. Bitter	0	1	2	3	4
12. Unhappy	0	1	2	3	4
13. Anxious	0	1	2	3	4
14. Worried	0	1	2	3	4
15. Energetic	0	1	2	3	4
16. Miserable	0	1	2	3	4
17. Muddled	0	1	2	3	4
18. Nervous	0	1	2	3	4
19. Angry	0	1	2	3	4
20. Active	0	1	2	3	4
21. Tired	0	1	2	3	4
22. Bad tempered	0	1	2	3	4
23. Alert	0	1	2	3	4
24. Uncertain	0	1	2	3	4