INTERNATIONAL JOURNAL OF METHODS IN PSYCHIATRIC RESEARCH, VOL. 5: 251–262 (1995)

DEVELOPMENT OF A SHORT QUESTIONNAIRE FOR USE IN EPIDEMIOLOGICAL STUDIES OF DEPRESSION IN CHILDREN AND ADOLESCENTS: FACTOR COMPOSITION AND STRUCTURE ACROSS DEVELOPMENT

STEPHEN C. MESSER* PhD, ADRIAN ANGOLD MRCPsych and E. JANE COSTELLO PhD
Developmental Epidemiology Program, Duke University Medical Center, Box 3454, Durham, North Carolina, USA

ROLF LOEBER PhD, WELMOET VAN KAMMEN PhD and MAGDA STOUTHAMER-LOEBER PhD
Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

SUMMARY

In the present study, the factor composition of a new, brief, easy-to-administer, depression inventory for children and adolescents, the Short Mood and Feelings Questionnaire (SMFQ), was examined. Using longitudinal data from the Pittsburgh Youth Study, confirmatory factor analyses were computed to evaluate the age-invariance of a single-factor SMFQ structure. Results provided strong support for the unidimensional structure of the SMFQ across grades 1–10 in a large sample of boys. Factor loading correlations across age groups were substantial suggesting developmental continuity in item (symptom) salience for the latent variable of depression. The magnitude of the mean factor loadings increased with age, suggesting that the depression construct was measured with less error as the children matured. In sum, the SMFQ appears to provide a brief and reliable measure of a core depression construct in children and adolescents.

KEYWORDS—depression; children and adolescents; assessment; factor analysis

INTRODUCTION

In a companion paper, we described the initial development of the Short Mood and Feelings Questionnaire (SMFQ) (Angold *et al.*, 1995). The aim was to produce a short, easy-to-administer, depression inventory for children and adolescents, tapping the central features of depressive symptomatology. Reviews of available childhood depression instruments and their psychometric properties have revealed a need for a brief, yet sound, epidemiological screening measure (e.g. Costello and Angold, 1988). Particularly for community epidemiological investigations where relatively low base rates of disorder are expected (e.g. 2–5%; Angold, 1988a, 1988b; Fleming and Offord, 1990), a brief screen or 'net' (Costello and Angold, 1988) with adequate operating characteristics would significantly increase the efficiency of such studies. The SMFQ, designed with this goal in mind, is a 13-item self-report depression checklist, with parallel child and parent versions. In our previous paper, the SMFQ was shown to: (a) correlate substantially with the Children's Depression Inventory (CDI) Kovacs, 1992) and the Diagnostic Interview Schedule for Children (DISC) (Costello *et al.*, 1982) depression scale; (b) discriminate depressed from non-depressed subjects in a general population sample; and (c) in exploratory factor analyses, show a unidimensional factor structure with high internal consistency (Angold *et al.*, 1995).

In designing the SMFQ, a guiding feature was to construct a single factor scale, including only items that loaded strongly on the first major factor extracted from a wider pool of items (i.e. the Mood and Feelings Questionnaire [MFQ] Costello and Angold, 1988). Understanding the factorial

*Author to whom correspondence should be addressed.

CCC 1049–8931/95/040251–12

Accepted 30 November 1994

composition of an instrument, like the SMFQ, is crucial in conducting analyses and deriving resultant generalizations. For example, to optimize the capacity of a measure to discriminate diagnostically distinct groups, the factorial structure of the instrument must be taken into consideration. A multidimensional depression inventory that is scored as a single scale is at the least missing an opportunity to capitalize on the richness of the measure to maximize validity and, at worst, providing misleading results. For example, studies examining the ability of the CDI to discriminate depressed from non-depressed psychiatric control groups have reached somewhat inconsistent results (e.g. Kazdin, 1988; Kovacs, 1992). These mixed findings may reflect, in part, the CDI's multidimensional structure, with at least two prominent factors representing dysphoria and oppositional behaviour (e.g. Weiss & Weisz, 1988).

In addition, understanding developmental shifts or invariance in the factor structure of our childhood depression instruments is essential. Such information highlights patterns of symptom covariation as a function of age. The limited available evidence points to a few age-graded changes in the phenomenology of symptoms comprising depressive disorder among school-aged and adolescent youth (e.g. Kovacs and Paulauskas, 1984; Mitchell *et al.*, 1988; Ryan *et al.*, 1987). Since depressive symptoms and disorders show an increasing prevalence with age, as well as a female preponderance emerging in adolescence (e.g. Anderson *et al.*, 1987; Cohen *et al.*, 1993; Angold, 1988a; Costello, 1990; Fleming and Offord, 1990; McGee *et al.*, 1992; Velez *et al.*, 1989), it is crucial to ascertain whether the patterns of SMFQ symptom covariation (i.e. factorial structure) vary with development. Finally, generalizations regarding developmental changes in depression will rest in part on the psychometric integrity and factorial composition of our depression measures. For example, asserting that there are changes in self – reported depressive symptomatology from early elementary to high school is based on the factorial unity of the depression measure. Demonstration of the relative developmental invariance of a robust depression factor would bolster such a generalization; otherwise, age-related differences could reflect changes in correlated syndromes such as oppositional or disruptive behaviour.

The purpose of this study, therefore, was to examine the factor composition and structure of the SMFQ. To achieve this aim, data were obtained from the Pittsburgh Youth Study, consisting of three cohorts of first-, fourth-, and seventh-grade boys, followed over 3–4 years, in an accelerated longitudinal design. Based upon the preliminary work with the SMFQ, the main hypothesis driving the current analysis was that, despite some variability in the magnitude and ranking of the SMFQ factor loadings across age groups, a single factor solution would most parsimoniously represent the longitudinal data. To facilitate the testing of this hypothesis, confirmatory factor analytic (CFA) model-fitting was applied (e.g. Joreskog and Sorbom, 1985) and various indices of the coherence of the SMFQ latent depression construct across age groups were examined (e.g. the rank order stability of the factor loadings, changes in the mean factor loading, total score stability).

METHOD

Subjects

The Pittsburgh Youth Study consists of three cohorts of boys who were in grades one, four and seven when the study began. Potential subjects were randomly selected from the list of all boys in these grades in the Pittsburgh school system. Of those selected, 84% of the families consented to participate, resulting in a sample of approximately 850 boys in each of the three grades. In the initial screening assessment, information on the boys' antisocial and prosocial behaviour was collected from caregivers, teachers and the boys themselves. A risk score was then calculated using the reports of potentially indictable offenses as a criterion. On the basis of the risk score, a final sample was selected for intensive assessments, consisting of the top 30% ($n = 250$) of the most antisocial boys in each grade and an equal number of boys selected randomly from the remainder. This sampling strategy resulted in three cohorts (first, fourth and seventh graders – i.e. youngest, middle and oldest samples) of about 500 boys each with whom follow-up interviews were conducted every 6 months for 4 years (except that the fourth graders were followed for only 3 years). The results reported here are based on SMFQ assessments conducted at entry into the study and the three follow-ups at yearly intervals

(hereinafter referred to as waves 1 to 4). Wave 1 assessments were performed on 1517 subjects; of these, seven had missing age or SMFQ data. Two boys were under 6 years of age and two were over 15, and due to their extreme ages in the context of the remaining sample and their small numbers, were dropped from the sample. Elimination of these 11 subjects resulted in a final wave 1 sample size of 1502. The same procedure resulted in a wave 2, 3 and 4 sample sizes of 1428, 1400 and 903, respectively (fourth graders were not retested in wave 4).

The subject characteristics, which have been previously described by Van Kammen, *et al.*, (1991), included the following: (a) about one-half of the boys were African-American and one-half Caucasian; (b) approximately 40% of the boys lived with a single parent; and (c) about 40% of the parents received public financial assistance.

Measure

The Short Mood and Feelings Questionnaire (SMFQ; Angold *et al.*, 1995) is a 13-item scale derived from a 34-item depression questionnaire (MFQ; Costello *et al.*, 1991). The SMFQ was designed to provide a rapidly administered checklist of core symptoms for children and adolescents ages 8–16 years. Each item is keyed to a three-point scale with the following possible responses: 'true' (scores 2), 'sometimes true' (scores 1) or 'not true' (scores 0). Children were assessed in their homes and completed the SMFQ as part of a package of self-report measures, or if necessary, had it read to them by the interviewer.

Data analysis

Confirmatory factor analyses (CFA) were performed using the LISREL VII program (Joreskog and Sorbom, 1985). CFA requires the specification of three parameter matrices: (1) *lambda*, containing the estimated factor loadings; (2) *phi*, includes the latent variable correlations, which in the case of a single factor solution, is not relevant; and (3) *theta delta*, containing the error (uniqueness, disturbance) terms. LISREL was constrained to a one-factor model with the 13 parameters representing the SMFQ item-factor correlations and their error variances (set free). Since we attempted to develop a unifactorial scale in the design stage of the SMFQ, we were concerned with testing whether a single-factor model did, in fact, fit new data adequately. Correlations between the disturbance terms were constrained (fixed) to zero. LISREL derives a reproduced correlation matrix based upon the three matrices and the fixed parameters specified by the user.

The LISREL program attempts to minimize a fit function (e.g. maximum likelihood) based on the difference between the original and reproduced data matrix. Among the output provided to assess the adequacy of the model tested are: (a) the goodness of fit index (GFI); (b) adjusted goodness of fit index (AGFI) (adjusted for sample size); (c) root mean square residual (RMSR) (an overall model error term); (d) efficient estimates of the free parameters; (e) a χ^2-based significance test for overall model fit; and (f) standard errors and significance tests for each fitted parameter. Since the χ^2 statistic is a function of sample size, one additional fit index was used, χ^2/degrees of freedom. Joreskog (1969) proposed this ratio as an index of model fit, particularly with large samples. Values below three are considered a conservative indicator of good model fit (Carmines and McIver, 1981). Following the recommendations of Hayduk (1987), each of these multiple indices will be evaluated to assess the model's adequacy.

Based upon earlier analyses of the SMFQ (Angold *et al.*, 1995), the distribution of SMFQ depression scores departed markedly from normality (as would be expected for a measure of psychopathology) showing a reversed-J-distribution. In addition, each item is based on a three-point ordinal scale. Joreskog and Sorbom (1985) suggested that when analyzing non-normal and/or ordinal data, the weighted least squares (WLS) fit function be used with an asymptotic covariance matrix generated from the polychoric correlations. Using WLS estimation is especially important for correct standard errors of parameter estimates and valid χ^2 goodness-of-fit measures (Joreskog and Sorbom, 1985). Therefore, in the confirmatory factor analyses, polychoric correlations with weight matrices were computed using PRELIS and input to LISREL.

The analysis plan followed several stages. First, tests were conducted for single-factor models with the wave-1 data for the first, fourth and seventh grade cohorts (i.e. youngest, middle and oldest samples). Second, as an additional 'check' on the appropriateness of a unifactorial model, exploratory factor analyses were computed for

Table 1a. Pittsburgh Youth Study: SMFQ confirmatory factor analyses by grade (cohort 1–youngest sample)

	Grade 1 ($n=495$)	Grade 2 ($n=478$)	Grade 3 ($n=477$)	Grade 4 ($n=470$)
		Factor loadings		
LISREL estimates	F1	F1	F1	F1
Q1 miserable	44	41	48	51
Q2 didn't enjoy	47	43	39	57
Q3 so tired	52	58	55	60
Q4 restless	57	58	55	64
Q5 no good	68	76	77	82
Q6 cried a lot	50	62	62	74
Q7 concentrate	58	58	57	75
Q8 hated self	72	79	87	87
Q9 bad person	77	78	84	83
Q10 lonely	53	67	68	73
Q11 nobody loves	70	73	82	81
Q12 never as good	72	73	77	81
Q13 everything wrong	68	78	83	90
Goodness of fit indices				
GFI	98	98	99	98
AGFI	97	98	98	97
RMS	07	06	05	07
χ^2/df	127/65	82/65	78/65	119/65
p	< 001	07	13	001
Coefficient alpha	81	–	–	–

Table 1b. Pittsburgh Youth Study: SMFQ confirmatory factor analyses by grade (Cohort II–middle sample)

	Grade 4 ($n=502$)	Grade 5 ($n=492$)	Grade 6 ($n=475$)
		Factor loadings	
LISREL estimates	F1	F1	F1
Q1 miserable	51	71	64
Q2 didn't enjoy	49	61	69
Q3 so tired	53	73	69
Q4 restless	45	63	70
Q5 no good	74	83	83
Q6 cried a lot	47	64	77
Q7 concentrate	65	73	73
Q8 hated self	75	83	90
Q9 bad person	79	83	89
Q10 lonely	70	77	86
Q11 nobody loves	81	83	88
Q12 never as good	80	74	83
Q13 everything wrong	80	84	93
Goodness of fit indices			
GFI	98	99	97
AGFI	97	98	96
RMS	06	05	09
χ^2/df	98/65	68/65	170/65
p	005	37	001
Coefficient alpha	84	–	–

Table 1c. Pittsburgh Youth Study: SMFQ confirmatory factor analyses by grade (Cohort III–oldest sample)

	Grade 7 ($n=505$)	Grade 8 ($n=458$)	Grade 9 ($n=448$)	Grade 10 ($n=433$)
	Factor loadings			
LISREL estimates	F1	F1	F1	F1
Q1 miserable	63	76	75	83
Q2 didn't enjoy	46	73	80	83
Q3 so tired	59	61	65	66
Q4 restless	63	66	68	73
Q5 no good	83	87	89	81
Q6 cried alot	75	77	66	75
Q7 concentrate	70	60	79	70
Q8 hated self	91	99	96	92
Q9 bad person	91	81	91	88
Q10 lonely	84	81	83	83
Q11 nobody loves	89	91	92	85
Q12 never as good	87	85	81	87
Q13 everything wrong	89	90	90	87
Goodness of fit indices				
GFI	98	98	98	97
AGFI	97	97	97	96
RMS	08	09	08	11
χ^2/df (<2)	132/65	124/65	121/65	162/65
p	<001	001	<001	<001
Coefficient alpha	87	–	–	88

F1, Factor one; GFI, goodness-of fit index; AGFI, adjusted goodness-of-fit index; RMS, root mean square residual; –, not computed.

these same three cohorts. To allow for the possibility of correlated common factors, principal axis factoring (PAF) was used with oblique (oblimin) rotation. Scree plots, simple structure and interpretability guided judgement of the number of factors. Third, when more than a single factor solution was suggested, the single-factor and n-factor models were statistically compared using LISREL and the χ^2 goodness-of-fit change test and the other indices of model adequacy, to determine the most parsimonious number of factors. Lastly, the 'best' solution generated by this process was replicated with the other waves of data.

RESULTS

SMFQ latent variable structure: initial confirmatory factor analyses for grades 1, 4 and 7

Results from the CFAs, conducted separately for subjects in the youngest, middle and oldest samples at wave 1, supported strongly the adequacy of a single- factor model (see Table 1). The various indices of fit each supported the adequacy of a single latent variable model across development: (a) goodness of fit (GFI) and adjusted goodness of fit (AGFI) indices were all high (≥ 0.97); (b) root mean square residual (RMSR) error terms were all low (≥ 0.08); and (c) χ^2/df indices were all less than 3. Therefore, as hypothesized, the SMFQ appeared to be a unidimensional scale, at least for the wave-1 data for the three cohorts.

To pursue the possibility of a multi-dimensional scale further, exploratory PAF analyses with oblique rotation were conducted on the SMFQ items for the same youngest, middle and oldest samples. The factor analyses revealed a consistent two-factor solution with the first and second factors explaining approximately 35 and 10% of the variance, and with eigenvalues around 5 and 1.5, respectively. However, closer examination of the second factor, typically comprised of items 1–4 and 7, revealed a factor consisting of items with

Table 2. Pittsburgh Youth Study: item score endorsement frequencies for first, fourth, seventh and tenth graders

Item score frequencies (%)	Grade 1 ($n = 505$)	Grade 4 ($n = 458$)	Grade 7 ($n = 448$)	Grade 10 ($n = 433$)
Q1 miserable				
true	27	25	07	07
sometimes true	38	36	29	28
not true	35	39	64	65
Q2 didn't enjoy				
true	16	13	06	04
sometimes true	29	24	23	24
not true	55	63	71	72
Q3 so tired				
true	20	19	14	08
sometimes true	27	31	32	45
not true	53	50	54	47
Q4 restless				
true	17	20	10	08
sometimes true	23	27	27	33
not true	60	53	63	59
Q5 no good				
true	12	06	02	02
sometimes true	16	13	06	08
not true	72	81	92	90
Q6 cried a lot				
true	18	06	00	00
sometimes true	26	14	05	03
not true	56	80	95	97
Q7 concentrate				
true	24	13	03	03
sometimes true	29	28	23	23
not true	47	59	74	74
Q8 hated self				
true	07	03	01	00
sometimes true	11	10	06	05
not true	82	87	93	95
Q9 bad person				
true	10	05	01	01
sometimes true	12	15	10	06
not true	78	80	89	93
Q10 lonely				
true	20	11	02	02
sometimes true	29	25	15	14
not true	51	64	83	84
Q11 nobody loves				
true	11	05	01	00
sometimes true	14	11	06	07
not true	75	84	93	93
Q12 never as good				
true	13	09	02	01
sometimes true	17	15	13	10
not true	70	76	85	89
Q13 everything wrong				
true	10	07	02	01
sometimes true	15	14	09	09
not true	75	79	89	90

Frequencies for grades 1, 4 and 7 reflect the wave-1 assessments for the three cohorts. The grade 10 data come from the wave-4 assessment of the third cohort.

Table 3. Twelve-month stabilities of the SMFQ

Grade	Stability
1	0.28
2	0.40
3	0.45
4	0.38
5	0.46
6	–
7	0.37
8	0.42
9	0.48

12-month stability data unavailable.

the highest frequencies (see Table 2). To assess whether a two-factor model provided a significantly better model fit, CFAs constrained to two factors (based on the PAFs) were calculated. Although χ^2 values dropped to a *statistically* significant degree according to nested tests (Hayduk, 1987), other indices of model fit (*i.e.* GFI, AGFI and RMSR) remained virtually unchanged, suggesting that the single-factor model should be retained.

The preceding analyses were conducted using the SMFQ item scores keyed in their original scales (i.e. 0, 1 or 2). As an additional test of the validity of the resultant model and since Schoenbach *et al.* (1982) found that, in comparison with adults, adolescents showed a response set characterized by a greater likelihood of endorsement of mild symptoms on the CES-D (but similar levels of assent to more persistent symptoms), CFAs were rerun to examine the relevance of this response style. Items were dichotomously recoded to capture more persistent symptomatology, i.e. 0 *or* 1 versus 2. As expected, the single-factor model continued to fit very well, with a consistent yet modest improvement in model fit with the positives limited to more persistent symptoms (i.e. 0 or 1 versus 2). Significant testing for model improvement was not possible since these models were not nested (Hayduk, 1987).

In sum, whether coding the SMFQ in its original format or with each item as a binary measure makes little difference in the factor structure of the scale. Additionally, two-factor models appear to reflect nothing but high and low endorsement frequencies, and the single-factor solution is clearly supported by the rule of parsimony. We then went on to fit single-factor CFAs to the data from waves 2–4 (see Table 1).

SMFQ latent variable structure: confirmatory factor analyses at three- or four-year follow-up

The various indices of fit each supported the adequacy of a single latent variable model across the entire age-range: (a) goodness of fit (GFI) and adjusted goodness of fit (AGFI) indices were all high ($\geqslant 0.96$); (b) root mean square residual (RMSR) error terms were all in the low range 0.05–0.11); and (c) χ^2/df indices were all less than 3 (see Table 1). These analyses also demonstrated that the factor structure was not changed substantially by repeated administrations.

Symptom salience: rank order consistency of factor loadings by grade

Having established the unifactorial composition of the SMFQ, we went on to explore the patterning of factor loadings and whether this pattern evidenced stability. Spearman correlations revealed significant and substantial correlations in the rank ordering of factor loadings (i.e. symptoms comprising the depression construct) between first and fourth graders ($r_s = 0.86$, $p < 0.001$), fourth and sixth graders ($r_s = 0.77$, $p < 0.001$) and seventh and tenth grade boys ($r_s = 0.80$, $p < 0.001$). Moreover, the factor loading correlation between first and tenth grade boys was large and significant ($r_s = 0.64$, $p < 0.01$). These data provided strong support for developmental continuity in the salience and patterning of the 13 depression symptoms in defining the latent variable of depression.

One-year stabilities of the SMFQ

To examine the long-term stability of the SMFQ, 12-month stability coefficients were calculated (see Table 3). These correlations ranged from 0.28 for first graders to 0.48 for boys in the ninth grade. Second graders showed a stability correlation of somewhat larger magnitude than the first graders (0.40), with the others ranging from 0.37 and 0.38 for seventh and fourth graders, to 0.45 and 0.46 for third and fifth graders, respectively.

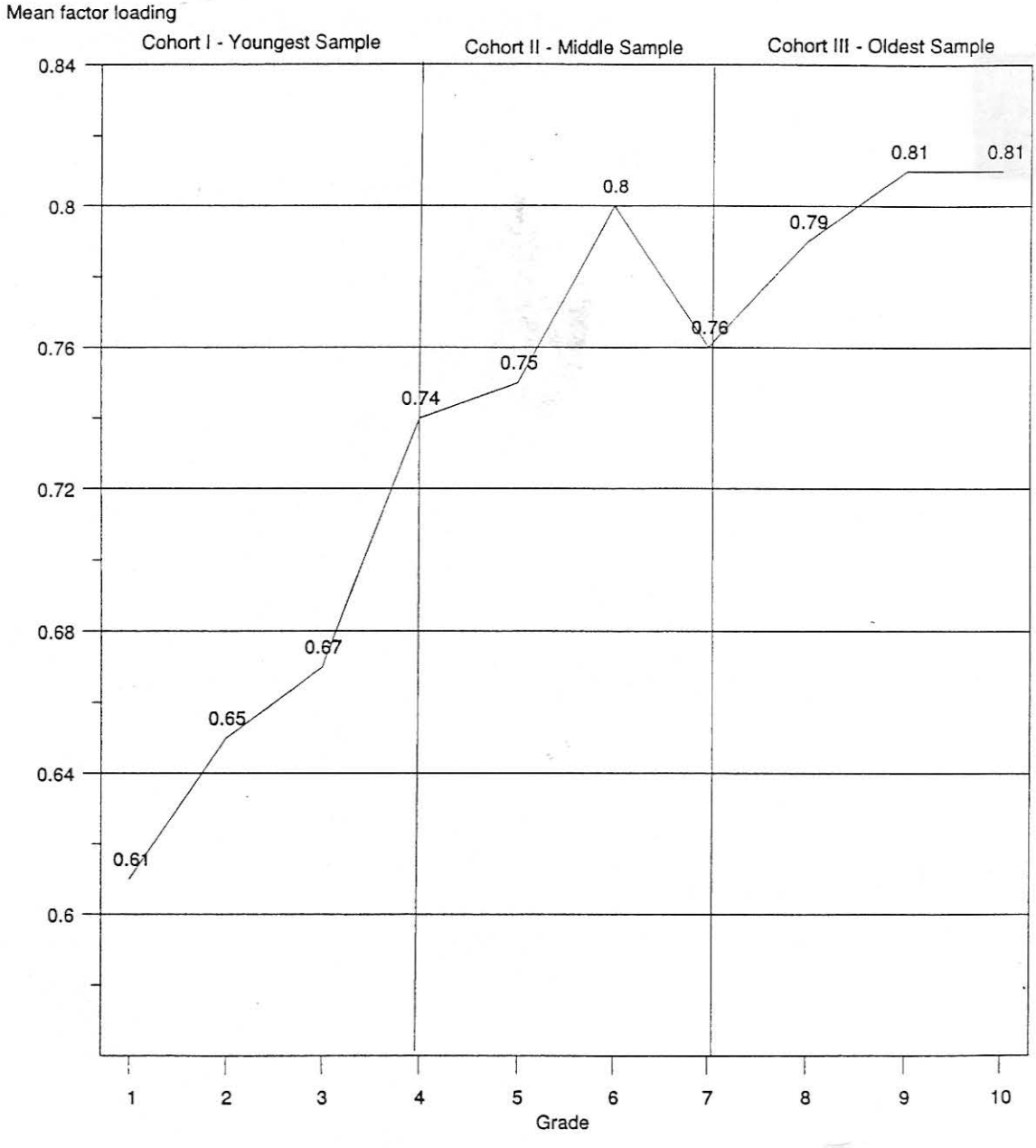

Fig. 1. SMFQ mean factor loadings by grade

Syndromal density or coherence: mean factor loadings by grade

Mean factor loadings can be interpreted as reflecting the ability of the latent depression construct to 'capture' the SMFQ symptom variance. In other words, higher mean loadings imply less residual error. Wilcoxon tests were used to compare the distributions and central tendencies of the factor loadings for the first and fourth, fourth and sixth, and seventh and tenth graders. Results revealed significant differences for first and fourth

graders ($Z = 3.2$, $p < 0.001$), and fourth and sixth graders ($Z = 3.2$, $p < 0.01$). No significant difference was found in the distribution of factor loadings for seventh and tenth grade youngsters ($p > 0.10$). These findings indicate that the mean factor loadings of the SMFQ items appear to increase with age from approximately 6 to 16 (see Fig. 1). Thus, the items comprising the latent depression variable, as assessed by the SMFQ, are measured with less error in older children.

Race differences in factor structure

To examine the comparability of the SMFQ's factor structure by race, separate CFAs were run for the white males ($n = 268$) and the African-American males ($n = 218$) using the middle cohort wave 1 (fourth grade) data. Single-factor solutions provided excellent fits for both race groups (white: χ^2 $(65) = 75.7$; $p = 0.17$, AGFI = 0.96, RMSR = 0.08; African-American: χ^2 $(65) = 67.7$, $p = 0.39$, AGFI = 0.97, RMSR = 0.07. Thus, the SMFQ's factor properties appeared similar for white and African–American boys.

DISCUSSION

In three large cohorts of randomly selected boys, the unifactorial composition of the SMFQ was supported. A single depression factor more parsimoniously explained the SFMQ item intercorrelations than a two-factor solution suggested by exploratory analyses. The second factor appeared to be a methodological artifact of high endorsement rates. Importantly, the relative age invariance of the depression construct, as measured by the SMFQ, was corroborated, as demonstrated by the replicability of the single-factor model with children and adolescents in grades 1–10.

The simple, unifactorial structure of the SMFQ provides a valuable asset for those interested in obtaining a quick and valid measure of core depressive symptomatology. Other available childhood depression instruments present a more complicated picture as to their factor compositions. For example, the factor structure of the CDI has been most extensively studied (e.g. Carey *et al.*, 1987; Helsel and Matson, 1984; Hodges *et al.*, 1982; Kovacs, 1983, 1992; Saylor *et al.*, 1984; Weiss and Weisz, 1988; Weiss *et al.*, 1992). In non-referred and clinical samples, factor analyses have suggested anywhere from two (Carey *et al.*, 1987; Hodges *et al.*, 1982) to eight factors (Saylor *et al.*, 1984). The CDI manual (Kovacs, 1992) proposes a five-factor solution based on analyses of the Finch *et al.* (1985) normative data for 1266 second through eighth-graders. These five factors were labeled negative mood, interpersonal problems, ineffectiveness, anhedonia and negative self-esteem. These divergent findings for the CDI factor analysis studies probably result from several sources: (a) the heterogeneity of the number and characteristics of subjects (e.g. age, gender, racial and diagnostic composition); (b) differing factor analytic methods and criteria (e.g. principal components with varimax rotation versus maximum likelihood with oblique rotation); and (c) the well-known subjectivity involved in proposing the best-fitting factor analytic model. Despite this range, all these studies identified a first major factor that accounts for the bulk of the variance among the CDI items (e.g. Kovacs, 1992; Weiss and Weisz, 1988; Weiss *et al.*, 1992). Notably, a 10-item CDI short-form has been recently developed (Kovacs, 1992) although no factor analytic data have been reported.

The demonstration of the age invariance of the SMFQ factor structure is one of the first empirical displays of such continuity. Only three investigations were identified that explicitly addressed the issue of developmental stability in the factor composition and structure of self-reported depression based on questionnaire responses. Roberts *et al.*, (1990) conducted a large-scale evaluation of the operating characteristics, including factor structure, of the Center for Epitemiological Studies–Depression (CES-D) scale (Radloff, 1977) in four samples of high school students. Multiple sample confirmatory factor analyses supported the invariance of the factor loadings across the adolescent age groups. Weiss and Weisz (1988) factor analyzed the CDI responses of 110 clinic-referred children and 139 adolescents. Three factor solutions were judged to best fit both the child and adolescent samples. Weisz *et al.* (1992) pooled data from several of the previously mentioned studies (Carey *et al.*, 1987; Nelson *et al.*, 1987; Saylor *et al.*, 1984; Weiss and Weisz, 1988; Weiss *et al.*, 1992), resulting in a sample of 1030 clinic-referred child and adolescent subjects. In the analyses most pertinent to the current investigation, the investigators tested the similarities of the factor

loadings of the child and adolescent CDIs by forcing a maximum liklihood factor analysis to a single solution for both samples. Results indicated that 10 of the factor loadings differed significantly between the children and adolescents, although with a type I error correction for multiple comparisons and large sample size, the number of differences dropped to three. Thus, despite differences in subject characteristics (e.g. community versus clinical) and methods, these studies are generally consistent with the current findings of child and adolescent age invariance in the factor characteristics of self-reported depression measures. Future studies of the developmental properties of construct assessments should capitalize on the strengths of the confirmatory factor analytic approach, with the additional advantage of facilitating comparisons across such studies.

In the present study, additional analyses investigated the continuity of the SMFQ's measurement properties. One-year rank order stability coefficients ranged from 0.28 to 0.48. In analyses of the age-graded rankings of symptoms defining the SMFQ depression construct, correlations of factor loadings were high, even from first to tenth grade, generating strong support for the continuity of the patterning of self-reported depressive symptomatology. Interestingly, the mean factor loadings increased with age, appearing to reach a stable peak around 10–12 years. This increased coherence was apparently not simply a function of increasing stability of SMFQ depression, since 12-month stability coefficients were comparable from second grade and up (ranging from 0.38 to 0.48). These findings are consistent with both community and clinical studies reporting moderate continuity of depressive symptoms and disorder (e.g. Asarnow *et al.*, 1988; Harrington *et al.*, 1990; Kovacs *et al.*, 1984; McGee and Williams, 1988). However, investigations into the varying levels of continuity of depressive symptomatology, syndrome, diagnosis and disorder, as well as associated antecedents and consequences of such stability, are areas with obvious clinical and public health importance. Lastly, further basic, construct validity work is necessary in the area of self-reported depressive symptomatology in general (e.g. Messer and Gross, 1995), since it has not been established to what degree that these measures are simply tapping a broader construct of negative affectivity (e.g. Stark *et al.*, 1991). Negative affectivity, as a personality or dispositional attribute, is probably embedded in a quite different network of developmental causal complexes and sequelae.

Two interpretations of the emerging development coherence of self-reported depression can be offered. First, from a methodological perspective, larger factor loadings represent more item variance explained by the factor. Thus, the somewhat smaller mean factor loadings for younger children may reflect more error variance. This 'error' could, for example, be due to younger children's more variable reading and comprehension levels. Or, younger children may be more easily distracted during testing, thus resulting in more error and hence smaller loadings. Substantively, the increasing sophistication of the child's representational abilities may underlie the observed pattern. In particular, advanced social-cognitive and self-conception development, characteristic of transitions from concrete operational to formal operational thought, may supply the child with the cognitive skills for the generation and maintenance of a complex representational schema such as a 'depressed self'. Relatedly, the child's understanding and attributions of the symptoms tapped by the SMFQ may become more refined with age. This would be manifested by less error (i.e. higher factor loadings) associated with the ratings of symptoms comprising the SMFQ checklist. The development of an increasingly coherent and 'tight' depression construct may indicate the emergence of a distinct psychopathological condition, such as more adult-like forms of depressive conditions. Whatever the explanations(s) for the changing magnitudes of factor loadings, it is important to note the strength of the single-factor solution across development and the overall similarities that overshadow the observed differences.

In summary, the SMFQ provides a brief, unifactorial measure of self-reported child and adolescent depression. The major limitation of the present study was the exclusive reliance on a predominantly high risk male sample. However, offsetting this limitation are the advantages of having three age cohorts with four annual follow-ups. Whether the unifactorial and age-invariant properties of the SMFQ will be found in female samples, particularly during the developmental period of dramatically rising depression prevalence rates, are questions addressable in another ongoing epidemiological investigation (Costello *et al.*, 1995).

ACKNOWLEDGEMENTS

Dr Messer was supported by a NIMH Post-Doctoral Fellowship Training Grant in Mental Health Services and Systems Research, MH-19117. Drs Messer, Angold and Costello were supported in part by the Leon Lowenstein Foundation and the William T. Grant Foundation (Angold and Costello). Data collection and the contributions of Drs Loeber, Van Kammen and Southamer-Loeber were supported by Grant 86-JN-CX-0009 from the Department of Justice. Points of view or opinions in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice.

REFERENCES

Anderson, J. C., Williams, S., McGee, R. and Silva, P.A. (1987) DSM–III disorders in preadolescent children: prevalence in a large sample from the general population. *Arch. Gen. Psychiat.* **152,** 69–77.

Angold, A. (1988a) Childhood and adolescent depression I: epidemiological and aetiological aspects. *Br. J. Psychiat.* **44,** 601–617.

Angold, A. (1988b) Childhood and adolescent depression II: research in clinical populations. *Br. J. Psychiat.* **153,** 476–492.

Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F. and Silver, D. (1995) Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int. J. Methods Psychiat. Res.* **5,** 237–249.

Asarnow, J. R., Goldstein, M. J., Carlson, G. A., Perdue, S., Bates, S. and Keller, J. (1988) Childhood-onset depressive disorders. A follow-up study of rates of rehospitalization and out-of-home placement among child psychiatric inpatients. *J. Affect. Disord.* **15,** 245–253.

Carey, M. P., Faulstich, M. E., Gresham, F. M., Ruggiero, F. M. and Enyart, P. (1987) Children's Depression Inventory: construct and discriminant validity across clinical and nonreferred (control) populations. Special Issue: Eating disorders. *J. Consult. Clin. Psychol.* **55,** 755–761.

Carmines, E. and McIver, J. (1981) Analyzing models with unobserved variables: analysis of covariance structures. In *Social Measurement: Current Issues* (G. Bohrnstedt and E. Borgatta, eds), pp. 65–115. Sage, Beverly Hills.

Cohen, P., Cohen, J. and Brook, J. (1993) An epidemiological study of disorders in late childhood and adolescence: 2. Persistence of disorders. *J. Child Psychol. Psychiat.* **34,** 869–877.

Costello, E. J. (1990) Child psychiatric epidemiology: Implications for clinical research and practice. In *Advances in Clinical Child Psychology* Volume 13, (B. B. Lahey and A. E. Kazdin, eds), pp. 53–90. Plenum Press, New York.

Costello, E. J. and Angold, A. (1988) Scales to assess child and adolescent depression: checklist, screens and nets. *J. Am. Acad. Child Adolescent Psychiat.* **27,** 726–737.

Costello, A. J., Edelbrock, C., Kalas, R., Kessler, M. D. and Klaric, S. H. (1982) *The National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC).* National Institute of Mental Health, Rockville, MD.

Costello, E. J., Benjamin, R., Angold, A. and Silver, D. (1991) Mood variability in adolescents: a study of depressed, nondepressed and comorbid paitents. *J. Affect. Disord.* **23,** 199–212.

Costello, E. J., Angold, A., Burns, N. J., Stangl, D., Tweed, D. and Erkanli, A. (1995) *The Great Smoky Mountains Study of Youth: I. Design, methods, and prevalence of DSM–III–R.* Manuscript submitted for publication.

Finch, A. J., Saylor, C. F. and Edwards, G. L. (1985) Children's Depression Inventory: sex and grade norms for normal children. *J. Consult. Clin. Psychol.* **53,** 424–425.

Fleming, J. E. and Offord, D. R. (1990) Epidemiology of childhood depressive disorders: a critical review. *J. Am. Acad. Child Adolescent Psychiat.* **29,** 571–580.

Harrington, R. C., Fudge, H., Rutter, M., Pickles, A. and Hill, J. (1990) Adult outcomes of childhood and adolescent depression: I. Psychiatric status. *Arch. Gen. Psychiat.* **47,** 465–473.

Hayduk, L. A. (1987) *Structural Equation Modeling with LISREL.* The Johns Hopkins University Press, Baltimore.

Helsel, W. J. and Matson, J. L. (1984) The assessment of depression in children: the internal structure of the Child Depression Inventory (CDI). *Behav. Res. Ther.* **22,** 289–298.

Hodges, K., Kline, J., Stern, L., Cytryn, L. and McKnew, D. (1982) The development of a child assessment interview for research and clinical use. *J. Abnorm. Child Psychol.* **10,** 173–189.

Joreskog, K. G. (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34,** 183–202.

Joreskog, K. and Sorbom, D. (1985) *LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood.* Scientific Software, Morresville, IN.

Kazdin, A. E. (1988) Childhood depression. In *Behavioral Assessment of Childhood Disorder,* 2nd ed., (E. J. Mash and L. G. Terdal, eds), pp. 157–195. The Guilford Press, New York.

Kovacs, M. (1983) *The Children's Depression Inventory: A Self-Rated Depression Scale for School-Aged Youngsters.* University of Pittsburgh School of Medicine, (Unpublished manuscript).

Kovacs, M. (1992) *Children's Depression Inventory (CDI) Manual.* Multi-Health Systems, Inc., North Tonawanda, NY.

Kovacs, M. and Paulauskas, S. L. (1984) Developmental stage and the expression of depressive disorders in children: an empirical analysis. In *Childhood Depression, New Directions for Child Development,* no. 26, (D. Cicchetti and K. Schneider–Rosen, eds), pp. 59–80. Jossey–Bass, San Francisco.

Kovacs, M., Feinberg, T. L., Crouse-Novak, M. A., Paulauskas, S. L., Pollock, M. and Finkelstein, R. (1984) Depressive disorders in childhood: II. A longitudinal study of the risk for a subsequent major depression. *Arch. Gen. Psychiat.* **41,** 643–649.

McGee, R. and Williams, S. (1988) A longitudinal study of depression in nine-year-old children. *J. Am. Acad. Child Adolescent Psychiat.* **27,** 342–348.

McGee, R., Feehan, M., Williams, S. and Anderson, J. (1992) DSM–III disorders from age 11 to age 15 years. *J. Am. Acad. Child Adolescent Psychiat.* **31,** (1), 51–59.

Messer, S. C. and Gross, A. M. (1994) Childhood depression and aggression: A covariance structure analysis. *Behaviour, Research, Therapy* **32,** 663–677.

Mitchell, J., McCauley, E., Burke, P. M. and Moss, S. J. (1988) Phenomenology of depression in children and adolescents. *J. Am. Acad. Child Adolescent Psychiat.* **27,** 12–20.

Nelson, W. M., Politano, P. M., Finch, A. J., Wendel, N. and Mayhall, C. (1987) Children's depression inventory: normative data and utility with emotionally disturbed children. *J. Am. Acad. Child Adolescent Psychiat.* **26,** 43–48.

Radloff, L. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Measurement* **3,** 385–401.

Roberts, R. E., Andrews, J. A., Lewinsohn, P. M. and Hops, H. (1990) Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychol. Assessment* **2,** 122–128.

Ryan, N., Puig-Antich, J., Ambrosini, P. J., Rabinovich, H., Robinson, D., Nelson, B., Iyengar, S. and Twomey, J. (1987) The clinical picture of major depression in children and adolescents. *Arch. Gen. Psychiat.* **44,** 854–861.

Saylor, C. F., Finch, A. J., Spirito, A. and Bennett, B. (1984) The Children's Depression Inventory: a systematic evaluation of psychometric properties. *J. Consult. Clin. Psychol.* **52,** 955–967.

Schoenbach, V. J., Kaplan, B. H., Grimson, R. C. and Wagner, E. H. (1982) Use of a symptom scale to study the prevalence of a depressive syndrome in young adolescents. *Am. J. Epidemiol.* **116,** 791–800.

Stark, K. D., Kaslow, N. J. and Laurent, J. (1991) The assessment of depression in children: are we assessing depression or the broad-band construct of negative affectivity? *J. Emotional Behav. Disorders* **1,** 149–154.

Van Kammen, W. B., Loeber, R. and Stouthamer-Loeber, M. (1991) Substance use and its relationship to conduct problems and delinquency in young boys. *J. Youth Adolescence* **20,** 399–413.

Velez, C. N., Johnson, J. and Cohen, P. (1989). A longitudinal analysis of selected risk factors of childhood psychopathology. *J. Am. Acad. Child Adolescent Psychiat.* **28,** 861–864.

Weiss, B. and Weisz, J. R. (1988). Factor structure of self-reported depression: clinic-referred children versus adolescents. *J. Abnorm. Psychol.* **97,** 492–495.

Weiss, B., Weisz, J. R., Politano, M., Carey, M., Nelson, W. M. and Finch, A. J. (1992). Relations among self-reported depressive symptoms in clinic-referred children versus adolescents. *J. Abnorm. Psychol.* **101,** 391–397.