

Short measures of youth psychopathology: Psychometric properties of the Brief Problem Monitor (BPM) and Behavior and Feelings Survey (BFS) in a Norwegian clinical sample

Kristian Rognstad (✉ kristian.rognstad@r-bup.no)

1Center for Child and Adolescent Mental Health, Eastern and Southern Norway

Siri Saugstad Helland

1Center for Child and Adolescent Mental Health, Eastern and Southern Norway

Simon-Peter Neumer

1Center for Child and Adolescent Mental Health, Eastern and Southern Norway

Silje Baardstu

3Norwegian Institute of Public Health

John Kjøbli

1Center for Child and Adolescent Mental Health, Eastern and Southern Norway

Research Article

Keywords:

Posted Date: February 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1315138/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Tracking clinical outcomes during therapy can be useful for improving both clinical practice and research. For repeated data collection, short, reliable, and valid measures of central aspects of psychopathology are necessary. The current paper investigates the psychometric properties of two short surveys for measuring central dimensions of psychopathology in youth.

Methods: We investigated factor structure and validity of the Norwegian translations of the Behavior and Feelings Survey (BFS) and the Brief Problem Monitor (BPM). The BFS has previously shown a two-factor structure and indications of validity as a measure of internalizing and externalizing problems in youth. Likewise, the BPM has support for a three-factor structure of internalizing, externalizing and attention problems. In our sample of 503 youths and caregivers in a Norwegian outpatient clinic, we conducted confirmatory factor analyses to test the assumed measurement models and further considered concurrent validity of the measures.

Results: The results suggest that the assumed measurement models for both questionnaires only partly fit our data but that subscales of the BFS and BPM still indicate convergent validity.

Conclusions: Alternative measurement models, and the usefulness and limitations of these short form questionnaires for internalizing and externalizing problems, are discussed.

Introduction

Psychotherapy is generally helpful for youth with mental health problems (Weisz et al., 2017), but a significant group do not seem to experience symptom relief (Warren et al., 2010). A development within psychotherapy that attempts to address this issue is the increased emphasis on outcome measurement and measurement-based care. Measurement feedback systems (MFS) use self-reports or other data sources to continuously collect information about patients in therapy. Such data can inform treatment decisions made by the therapist throughout the course of therapy. In this sense, patient-reported outcome measures can potentially increase treatment effects as indicated by several reviews (Kendrick et al., 2016; Knaup et al., 2009; Shimokawa et al., 2010).

Prerequisites for a useful MFS are that the included measures have sound psychometric properties and that data collection can be implemented. When measurement is conducted repeatedly over the span of a therapy period, the number of items is highly relevant. Shorter forms are less taxing for respondents and are likely to increase response rates. Brief surveys are also preferable in research to ensure high response rates and attention to items.

Additionally, feedback should be of clinical relevance and the measures thus gauge concepts that are appropriate for the patient population. On this background, the aim of this study is to evaluate two short measurements for central dimensions of psychopathology.

A two-dimensional model for the classification of child and adolescent psychopathologies—internalizing and externalizing problems—is often regarded as a parsimonious and adequately precise way to capture child and adolescent disorders (Achenbach & Edelbrock, 1981; Krueger et al., 1998). The internalizing dimension encompasses anxiety and depression symptoms, while the externalizing dimension covers aggressive, delinquent, and hyperactive-impulsive behavior (Krueger et al., 1998). Thus, reporting on patients' fluctuation on these dimensions can be useful for treatment and research.

The Brief Problem Monitor (BPM)

One frequently used set of measures in youth populations is the Achenbach System of Empirically Based Assessment (ASEBA) forms, which includes the Child Behavior Checklist (CBCL) and the Youth Self-Report (YSR) (Achenbach & Rescorla, 2001). The Brief Problem Monitor Parent (BPM-P) and Youth (BPM-Y) forms are short versions of the CBCL and the YSR, respectively. These shorter forms are designed to measure the three factors of internalizing, externalizing and attention problems (Achenbach et al., 2011).

The BPM has been developed through factor analysis and item response theory (IRT) for a large archival CBCL dataset (Achenbach, 2011). Different studies have provided evidence that the BPM meets some important psychometric standards. In Piper et al. (2014) the BPM-P had good internal consistency for the full scale (Cronbach's alpha = .91) and the subscales (Cronbach's alpha = .78 - .87) and excellent correspondence with both the CBCL total score ($R^2 = .90$) and its subscales (internalizing $R^2 = .74$, externalizing $R^2 = .86$, and attention problems $R^2 = .94$). The Norwegian translation of the BPM has been found to have acceptable reliability and correspondence with the long version of ASEBA to indicate validity (Richter, 2015). In a more recent study, BPM-P had excellent model fit for the three-factor model in a sample of 8–12-year-old children identified as being at risk for internalizing problems (Pedersen et al., 2021).

The Behavior and Feelings Survey (BFS)

The Behavior and Feelings Survey (BFS) is a 12-item survey that measures internalizing and externalizing problems in children and adolescents (Weisz et al., 2019). The BFS is very brief and thus minimizes the measurement burden. In contrast to the BPM, the BFS is freely available (<https://weizslab.fas.harvard.edu/measures>), which eliminates financial and copyright barriers.

The BFS was developed and evaluated in four samples of youths, and the original English version has demonstrated a robust factor structure, good internal consistency, test-retest reliability, and convergent and discriminant validity (Weisz et al., 2019). The scale was developed in several steps: 1) youth and caregivers generated “top problems” and, based on these, an expert panel selected 48 items; 2) youths and caregivers answered the survey, and the number of items was reduced using IRT; 3) two latent factors were identified through exploratory factor analysis and each factor was reduced to 6 items using IRT; 4) the validity of 12-items BFS version was tested in a new sample; and 5) its sensitivity to change

was tested to consider its performance as a progress monitoring tool. In Weisz et al. (2019), the BFS met accepted psychometric standards and was shown to be sensitive to change during treatment. It had good internal consistency for both caregiver and youth reports for the total score (respectively $\alpha = .87$, and $\alpha = .87$) and for both the internalizing ($\alpha = .84$, and $\alpha = .91$) and the externalizing ($\alpha = .94$, and $\alpha = .89$) scales. Associations between BFS total and CBCL/YSR total scores were high for both caregiver ($r = .61$, $p < .001$) and youth ($r = .72$, $p < .001$) reports, indicating convergent validity.

Both BPM and BFS measurement systems include youth and caregiver versions, which is important as otherwise reliable measures have revealed modest cross-informant correlations, indicating that multiple perspectives are necessary (Rescorla et al., 2013; Achenbach, 2006; De Los Reyes & Kazdin, 2005).

Associations between scales and diagnoses

Some categories of the International Classification of Diseases (ICD) (World Health Organization, 2002) contain mainly internalizing or externalizing symptoms. Measures of internalizing and externalizing problems should thus have some relation to the diagnoses young people are given. Some studies have found links between groups of patients with different diagnoses and their scores on ASEBA CBCL/YSR and BPM. Associations have been established between the internalizing scale of ASEBA (CBCL/YSR) and general diagnostic groups (e.g. anxiety and affective disorders) as well as with more specific diagnostic groups (e.g., conduct disorder) (Kasius et al., 1997). In Ebesutani et al. (2019), CBCL scales corresponded to the diagnoses they targeted, although the prediction accuracy varied from not significantly better than chance (for the Aggressive behavior scale) to fair (e.g., Withdrawn/Depression scale) or good (e.g., Anxiety/Depressed scale). Using the shorter BPM version, Piper et al. (2014) reported that patients with a depression diagnosis had scores on the internalizing scale that were 3.6 times higher than those reporting no diagnosis, and patients with an anxiety diagnosis had internalizing scores 3.2 times higher. To our knowledge, our study is the first study to tie BFS scores to patient diagnoses.

In the current study, we were interested in whether the BPM and BFS could remain reliable and valid in Norwegian and in a Norwegian specialist health service context. The purpose of the study was to 1) test the BFS and BPM factor structure, 2) examine reliability, and 3) evaluate convergent validity of the translated BFS and BPM when used in a youth outpatient population in Norway.

Methods

Participants

503 patients were recruited from an outpatient clinic in Norway. Respondents were both patients ($n = 386$) and caregivers ($n = 412$). The patients were between 6 and 18 years old (mean age = 12.73, SD = 3.15) and 56% female. The clinic is part of the Norwegian specialist health service where should have one or more diagnoses registered, and in our sample the most frequent were anxiety disorders (35%), depression and mood disorders (22%), and behavioral and attention disorders (22%).

Measures

The Brief Problem Monitor has parallel forms for self-report for youth (BPM-Y), caregiver reports (BPM-P), and for teacher reports (BPM-T). The 19-item form is designed to measure three factors: seven items measuring externalizing problems (e.g., “I argue a lot” and “I disobey my parents”), six items measuring internalizing problems (e.g., “I feel worthless or inferior” and “I am too fearful and anxious”) and six items measuring attention difficulties (e.g., “I fail to finish things I start” and “I have trouble sitting still”).

The Behavior and Feelings Survey (BFS) consists of 12 items. Each item is rated from 0 (not a problem) to 4 (a very big problem) over the past week. Six items are intended to measure internalizing problems (e.g., “I feel sad” and “I worry about bad things happening”) and six items to measure externalizing problems (e.g., “I refuse to do what adults tell me to do” and “I argue with people”). The BFS has parallel self-report and caregiver forms.

The diagnoses followed the diagnostic system of the ICD 10th Revision (ICD-10). Multiple diagnoses could be given and in the current dataset patients had up to four diagnoses. The research team gave no instructions for the process of diagnosis and diagnoses in the dataset were extracted from patient journals after therapists followed ordinary procedures at the clinic.

Procedure

The data collection and processing were approved by the Regional Committee for Medical and Health Research Ethics – South East and further guidance for consent, privacy, and data protection were given by the Norwegian Center for Research Data. Patients were informed about the study through a letter before their first session. Upon arrival for the session, they were given a tablet with study information and, if consenting, filled out questionnaires on the tablet. Participants were instructed to fill out the forms independently, but younger patients could have items read or explained to them if necessary.

The Norwegian versions of both the BFS and BPM used were translated by Norwegian expert committees of researchers and psychologists and back-translated and reviewed by the original developers in accordance with WHO recommendations for the translation and adaptation of instruments (WHO, 2016).

Analytic Strategy

Reliability scores, correlation calculations and linear regressions were performed using IBM SPSS Statistics v. 27 (IBM Corp. 2020). Confirmatory factor analyses were conducted using Mplus 8.6. (Muthen & Muthen, 1998-2017).

Reliability. For both BPM and BFS, Cronbach’s alphas were used to measure internal consistency for subscales and total score, for both caregiver and youth versions. Internal consistency can be considered excellent when Cronbach’s alpha is ≥ 0.80 , good when $0.70 \leq r < 0.80$, adequate when $0.60 \leq r < 0.70$

and inadequate when < 0.60 (European Federation of Psychologists' Association, 2013). As a measure of inter-rater reliability, intraclass coefficients (ICC; two-way random model with measures of consistency) were calculated to assess agreement across youth and caregiver reports. In line with Koo & Li (2016), we based our evaluation of the ICC values on a 95% confidence interval of the ICC estimates and considered values < 0.5 as poor, 0.5-0.75 as moderate, 0.75-0.90 as good and > 0.90 as excellent.

Confirmatory Factor Analysis. We performed confirmatory factor analyses (CFA) to examine the BPMs intended three-factor structure of internalizing, externalizing and attention problems (Achenbach, 2011). Informed by the exploratory factor analysis (EFA) in Weisz et al. (2019), we conducted a CFA of the BFS to consider whether the same factor structure was reproduced in our sample using the Norwegian translation. All CFAs were conducted for youth and caregiver data separately.

Variables were treated as categorical as items were not assumed to have the same interval between each response alternative. The data from BFS and BPM total scores also failed tests of normality (both Kolmogorov-Smirnov and Shapiro-Wilk tests). Thus, a weighted least square mean and variance adjusted (WLSMV) estimation was applied (Li, 2016). Factor loadings were examined to determine whether they adhered to the assumed model, with substantial loadings on expected factors and no cross-loadings.

Given the large sample size, chi-square test for model fit are not eligible as this fit index might lead to rejection of reasonably well-specified models. Thus, several other fit indices were considered, including the root mean square error of approximation (RMSEA), the Tucker-Lewis index (TLI), the comparative fit index (CFI) and the Standardized Root Mean Squared Residual (SRMR). CFI and TLI values of $>.95$ indicate an acceptable model fit (Hu & Bentler, 1999), RMSEA $< .08$ indicates a fair fit (Browne & Cudeck, 1993), and approximate fit is defensible with SRMR ≤ 0.08 (Kline, 2016). Modification indices were consulted in cases of poor model fit.

Tests of validity. Convergent validity is an estimate of the measure's ability to agree with the outcome for another measure intended to measure the same construct. Concurrent convergent validity is good when test scores have a strong relationship with a measurement of the same construct administered at the same time or shortly after. In the current study, we assessed convergent validity by exploring the extent to which scores on the BFS are correlated with scores from the BPM. Pearson product-moment correlation coefficients were calculated for each of the BFS and BPM subscales' covariance with each other.

In addition, we considered whether patients with internalizing or externalizing diagnoses scored higher on internalizing or externalizing in the BFS/BPM subscales. Most participants in our sample had been given one or more diagnoses from ICD-10. Although the ICD-10 does not use the internalization/externalization dimensions as a theoretical framework, there should be some relation between scores on the subscales and a young person's diagnoses. Diagnoses were dummy coded for presence of internalizing diagnosis and presence of externalizing diagnosis (see appendix for an overview of categorization of diagnoses). We compared the scores on relevant subscales for patients with and without internalizing and externalizing diagnoses separately.

Results

Descriptive statistics

Mean scores and standard deviations for the BFS and BPM total scale and subscales are presented in Appendix 1.

Tests of Reliability

Internal reliability was good for the BFS youth (Cronbach's $\alpha = .84$), and for subscales of internalizing (Cronbach's $\alpha = .89$) and externalizing problems (Cronbach's $\alpha = .84$). For the BFS caregiver version, internal reliability was equally good (Cronbach's $\alpha = .85$), and excellent for subscales of internalizing (Cronbach's $\alpha = .90$) and externalizing (Cronbach's $\alpha = .91$) problems. Internal reliability was good for both the BPM-P (total Cronbach's $\alpha = .84$; internalizing Cronbach's $\alpha = .81$; externalizing Cronbach's $\alpha = .85$; attention problems Cronbach's $\alpha = .84$) and the BPM-Y (total Cronbach's $\alpha = .85$; internalizing Cronbach's $\alpha = .84$; externalizing Cronbach's $\alpha = .74$; attention problems Cronbach's $\alpha = .78$).

Inter-rater reliability, between patient and caregiver reports, was indicated by an interclass correlation coefficient of 0.64 ($p = .000$) for the BFS total score, 0.73 ($p = .000$) for the BFS internalizing subscale, and 0.69 ($p = .000$) for the BFS externalizing subscale. For the BPM, the inter-rater reliability was lower, with an interclass correlation coefficients of 0.52 ($p = .000$) for the BPM total, 0.65 ($p = .000$) for the internalizing scale, 0.67 ($p = .000$) for the externalizing scale, and 0.64 ($p = .000$) for the attention problems scale.

Factor structure

BPM factor structure. The three-factor model previously shown for the BPM was not confirmed in the current sample for neither the caregiver nor the youth report. For the BPM-P, chi square test for model fit was significant and none of the indices for approximate model fit were acceptable ($\chi^2 = 766.67$, $df = 149$, $RMSEA = .10$, $TLI = .88$, $CFI = .90$, $SRMR = .11$). BPM-Y was only on par with two of the four model fit indices in our analysis plan ($\chi^2 = 386.18$, $df = 149$, $RMSEA = .06$ and $SRMR = .08$, but with $CFI = .94$ and $TLI = .93$). Introducing a new factor by splitting internalizing into depression and anxiety items significantly increased model fit. This resulted in a four-factor model that performed well on all four investigated model fit indices for youth reports ($\chi^2 = 302.99$, $df = 146$, $RMSEA = .05$, $CFI = .96$, $TLI = .95$, $SRMR = .07$); however, model fit was still inadequate for the caregiver report data ($\chi^2 = 701.74$, $df = 146$, $RMSEA = .10$, $CFI = .91$, $TLI = .89$, $SRMR = .10$).

In this four-factor model, four items loaded on depression (in BPM-Y .67–.83, in BPM-P .61–.77), two items loaded on anxiety (in BPM-Y .87–.88, in BPM-P .80–.95), seven loaded on externalizing (in BPM-Y

.50–.79, in BPM-P .64–.87) and six loaded on attention problems (in BPM-Y .50–.89, in in BPM-P .53–.92).

All BPM subscales were highly correlated with the total score in both caregiver and youth reports ($r = .70 - .80$), except caregiver reports for internalizing, which were moderately correlated with total problems ($r = .52$). In the BPM subscales, there was a substantial correlation between externalizing and attention problems (caregiver $r = .53$, youth $r = .53$) but weak correlation between internalizing and externalizing (caregiver $r = .09$, youth $r = .33$), and internalizing and attention problems (caregiver $r = .03$, youth $r = .32$).

BFS factor structure. We attempted to reproduce Weisz et al.'s (2019) two-factor structure for both the caregiver and youth report versions of BFS through using CFA. Model fit for the two-factor model of the BFS was unsatisfactory (caregiver report: $\chi^2 = 544.88$, $df = 53$, $RMSEA = .14$, $CFI = .96$, $TLI = .95$, $SRMR = .07$, and youth report: $\chi^2 = 291.45$, $df = 53$, $RMSEA = .10$, $CFI = .97$, $TLI = .96$, $SRMR = .06$). As for the BPM, splitting the internalizing factor into a depression factor and an anxiety factor significantly improved model fit for both the caregiver and youth report data. In the BFS Youth, with a 3-factor model, the model fit was acceptable on all the fit indices considered ($\chi^2 = 192.39$, $df = 51$, $RMSEA = .08$, $CFI = .98$, $TLI = .97$, $SRMR = .05$). For the caregiver version, the three-factor model of externalizing, anxiety and depression obtained adequate model fit on three of the five fit indices ($\chi^2 = 347.33$, $df = 51$, $RMSEA = .11$, $CFI = .98$, $TLI = .97$, $SRMR = .05$).

For the caregiver report, three items loaded on a depression factor, with loadings between .80 and .96, three items loaded on an anxiety factor, with loadings between .85 and .92, and the remaining six items loaded between .81 and .90 on an externalizing latent factor. The same items in BFS youth loaded on similar latent factors, with loadings of .80–.95, .73–.90, and .69–.86, respectively.

Caregiver-reported internalizing and externalizing through the BFS were weakly correlated ($r = .15$, $p < .001$), but each of the subscales was highly correlated with BFS total problems, with internalizing-total $r = .78$, $p < .001$ and externalizing-total $r = .73$, $p < .001$. A similar pattern emerged in the youth-reported BFS, with internalizing-externalizing $r = .22$, $p < .001$, internalizing-total $r = .87$, $p < .001$, and externalizing-total $r = .67$, $p < .001$.

Convergent validity

Convergent validity analysis was performed by exploring the relationship between BFS and BPM scores. The correlations between the BFS subscales and the BPM subscales is presented in Table 1 (caregiver-report) and in Table 2 (youth-report).

The BFS total problem score was highly correlated with BPM total scores for both caregiver ($r = .69$, $p < .000$) and youth reports ($r = .77$, $p < .000$). Scores for BFS internalizing were highly correlated with BPM internalizing for both caregiver ($r = .74$, $p < .01$) and youth reports ($r = .83$, $p < .01$). Moreover, high correlations were obtained between BFS externalizing and BPM externalizing for caregiver ($r = .83$, $p < .000$) and youth reports ($r = .83$, $p < .000$).

0.01) and youth reports ($r = .76, p < 0.01$). Smaller correlations were seen between measures of internalizing and externalizing problems, and moderate associations were obtained between all measures of externalizing problems and BPM attention problems.

Table 1

Sum score correlations - caregiver report

	BFS int	BFS ext	BPM int	BPM ext	BPM att
BFS int	1.000	.148**	.744**	.117*	.013
BFS ext	.148**	1.000	-.016	.828**	.502**
BPM int	.744**	-.016	1.000	.087	.031
BPM ext	.117*	.828**	.087	1.000	.530**
BPM att	.013	.502**	.031	.530**	1.000

Correlations between measures of similar constructs in bold.

** . Significant at the 0.01 level (2-tailed).

* . Significant at the 0.05 level (2-tailed).

Table 2

Sum score correlations - youth report

	BFS int	BFS ext	BPM int	BPM ext	BPM att
BFS int	1.000	.215**	.834**	.171**	.250**
BFS ext	.215**	1.000	.193**	.763**	.496**
BPM int	.834**	.193**	1.000	.204**	.325**
BPM ext	.171**	.763**	.204**	1.000	.532**
BPM att	.250**	.496**	.325**	.532**	1.000

Correlations between measures of similar constructs in bold

** . Significant at the 0.01 level (2-tailed).

Relations between subscales and diagnostic groups

Overall, 256 participants had one or more internalizing diagnoses and 131 participants had one or more externalizing diagnoses. Fifty-six participants only had diagnoses that we did not code into internalizing

or externalizing problems, while 126 were missing diagnoses due to early drop out, early termination of therapy, or administrative errors.

Patients that were given diagnoses with either predominantly internalizing or externalizing difficulties had higher scores on the respective internalizing and externalizing subscales of BFS/BPM. The patients with an internalizing diagnosis reported a 1.9 times higher score on BFS internalizing self-report than the patients without such a diagnosis. Likewise, BFS internalizing caregiver-reported scores were 1.7 times higher for the patients with an internalizing diagnosis. BPM-Y internalizing scores were 1.8 times higher and BPM-P internalizing scores 1.5 times higher for patients with internalizing diagnoses than for those without. For patients with externalizing diagnoses, higher scores were obtained on BFS youth externalizing (1.7 times), on BFS caregiver externalizing (1.7 times), on BPM-Y externalizing (1.7 times) and on BPM-P externalizing (1.5 times) than for those without such diagnoses.

Discussion

Overall, results showed that the brief measures evaluated in this study, namely the BPM and the BFS, have many qualities that make them ideal for frequent monitoring of youth in therapy. They are eligible for use in measurement-based care by informing practitioners about fluctuations of central dimensions of mental health issues for young patients, prompting therapists to adjust therapy plans.

Weisz et al. (2019) found the BFS to be psychometrically sound with a two-factor structure and convergent and discriminant validity in relation to other well-established measures. However, this two-factor structure was not confirmed in our data using a Norwegian population and translation. Rather, we found support for a three-factor model reflective of anxiety, depression and externalizing problems which had loadings in an expectable range ($> .69$) and good model fit. Furthermore, indications of validity were also obtained as the subscales correlated with the relevant BPM subscales and were related to relevant diagnoses.

The BPM is the shortened version of the empirically well-established CBCL and YSR. From both US and Norwegian samples, a three-factor model for the BPM has been reported (Pedersen et al., 2021) along with indications of convergent validity (Piper et al., 2014; Richter, 2015). A limitation of these studies is that they investigate the BPM in the context of the entire CBCL/YSR and all analyses are based on taking the BPM items out of those measures. These items are then compared with the full scale. This approach, however, could influence the response set and inflate correlations as measurement error could be overlapping. In our sample, the three-factor model was not confirmed as a more complex model with four factors (anxiety, depression, externalizing and attention problems) had significantly better model fit. However, BPM total scores and subscales were related to the expected BFS subscales and diagnoses, indicating convergent validity.

For both measures, internal consistency was good to excellent for both subscales and total score in caregiver reports and self-reports. Inter-rater reliability was moderate (from 0.52 to 0.73) between self-reports and caregiver reports. This was expected based on previous studies that often find moderate to

weak inter-rater reliability with different type of informants (e.g., Rescorla et al., 2013; Achenbach, 2006) and illuminates the need for multi-informant approaches in mental health reporting.

Similar constructs measured in the BFS and BPM were highly correlated in all subscales and in caregiver and youth reports (from .74 to .83). As expected, moderate correlations were found between attention problems measured by the BPM and externalizing problems in the BFS (caregiver report .50, youth report .50) and the BPM (caregiver report .53 and youth report .53). Minor correlations were found across the different constructs of internalizing and externalizing problems.

The expected subscales were elevated among youth with diagnoses of internalizing and externalizing character. In both the BFS and the BPM, substantial and significant differences in scores on the internalizing scale were obtained for those with or without internalizing diagnoses. Similarly, patients with externalizing diagnoses had higher scores on externalizing scales than those without such diagnoses. This is in line with previous findings where higher BPM subscale scores were recorded among participants with theoretically related diagnoses (Piper et al., 2014). Larger differences were found in Piper et al. (2014), but this was not unexpected as their sample came from a general population. In our clinical sample, we expect more symptoms of internalizing problems also among those with externalizing diagnoses and vice versa. However, both the BFS and the BPM showed the same tendency in our sample.

Limitations

The current sample was composed entirely of patients who had been referred to the specialist health services and were considered by the treating clinic to fulfill the criteria of at least one ICD diagnosis. Thus, the present study may be influenced by selection factors associated with referrals and evaluations of whether clients should receive treatment. Although BFS and BPM are constructed as clinical tools, it would be of interest to see similar evaluations in non-clinical populations or in populations with less severe problems.

To get acceptable model fit, one of the factors was split into two. Splitting the internalizing factor into depression and anxiety significantly increased model fit for both the BFS and the BPM in caregiver and youth reports. Although anxiety and depression are both part of the internalizing problems construct reproduced in numerous studies, a measurement of very few items might have difficulties finding the overlapping aspects between them. Another possibility is that this is an artifact of translation.

The validity checks are limited to correlations between the investigated measures and between the measures and groups of diagnoses. Diagnoses are not distinct indicators of internalizing and externalizing problem factors as there are no theoretical implications in the ICD. Nor can we provide reliability scores for diagnoses as the data rely on a naturalistic diagnostic procedure. Alternatively, diagnoses could have been mapped more systematically, and we suspect that our method of extracting diagnoses from medical journals is likely to underestimate comorbidity. Another limitation with respect to interpretation of the link between diagnoses and scores on the BFS/BPM is that therapists were not blinded to the results of the BFS/BPM when making diagnoses.

Conclusion

The BFS and the BPM are short, easy to use, and have acceptable psychometrics for most purposes. Both measure general dimensions in psychopathology and can be useful in both MFS and in research with repeated measures. They are easily interpreted, with scores representing total problems and well-known and thoroughly documented factors in youth psychopathology. The parallel forms allow for patient and caregiver perspectives and hence triangulation and a better understanding of cases. BFS is also free to use which makes it easily accessible.

Declarations

Funding:

No external funding

Authors' contributions:

KR, SN and JK planned the study in collaboration with a project group at the clinic. KR oversaw data collection and prepared data files for analysis in collaboration with Einar Amlie. KR had full access to all the data in the study and takes responsibility for the integrity of the data in the study and the accuracy of the data analysis. Analyses were performed by KR with guidance from SSH and SB. KR wrote drafts and final version, while all authors reviewed drafts. All authors reviewed and accepted the final version.

Acknowledgements:

We thank our collaborators at Lovisenberg Diaconal Hospital and Nic Waals Institutt and the participants of the study.

References

1. Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15*(2), 94-98.
2. Achenbach, T. M., & Edelbrock, C. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged 4 through 16. *Monographs of the Society for Research in Child Development, 46*(188).
3. Achenbach, T. M., McConaughy, S., Ivanova, M. & Rescorla, L. (2011). *Manual for the ASEBA brief problem monitor (BPM)*. Burlington, VT: ASEBA
4. Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

5. Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A.c Bollen and J. S. Long (Eds.), *Testing structural equation models (pp. 136-162).*c Newbury Park, CA: Sage.
6. Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., ... & Moffitt, T. E. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin birth cohort study. *JAMA network open*, 3(4), e203221-e203221.
7. De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483-509.
8. Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillan, C. K., & Weisz, J. R. (2010). Concurrent validity of the Child Behavior Checklist DSM -oriented scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment*, 32(3), 373-384.
9. European Federation of Psychologists' Association (2013). *EFPA Review Model for the description and evaluation of psychological and educational tests.* (<http://assessment.efpa.eu/documents/> Accessed 09.09.2021.)
10. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
11. IBM Corp. Released 2020. *IBM SPSS Statistics for Windows, Version 27.0.* Armonk, NY: IBM Corp.
12. Kasius, M. C., Ferdinand, R. F., van den Berg, H., & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 38(6), 625-632.
13. Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., ... & Moore, M. (2016). Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database of Systematic Reviews*, (7).
14. Kline, R. B. (2016). *Principles and practice of structural equation modeling (4th ed.)*. New York: Guilford.
15. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163.
16. Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *The British Journal of Psychiatry*, 195(1), 15-22.
17. Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM III-R): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*.107, 216–227.
18. Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949.

19. Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus user's guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
20. Pedersen, M.L., Jozefiak, T., Sund, A.M. et al. (2021). Psychometric properties of the Brief Problem Monitor (BPM) in children with internalizing symptoms: Examining baseline data from a national randomized controlled intervention study. *BMC Psychology* 9, 185. <https://doi.org/10.1186/s40359-021-00689-1>
21. Piper, B. J., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of brief problem monitor, an abbreviated form of the child behavior checklist. *Psychiatry and Clinical Neurosciences*, 68(10), 759-767.
22. Rescorla, L. A., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ... & Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*, 42(2), 262-273.
23. Richter, J. (2015) Preliminary evidence for good psychometric properties of the Norwegian version of the Brief Problems Monitor (BPM), *Nordic Journal of Psychiatry*, 69(3), 174-178, DOI: 10.3109/08039488.2014.951070
24. Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298.
25. Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*, 78(2), 144.
26. Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., ... & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. *American Psychologist*, 72(2), 79.
27. Weisz, J. R., Vaughn-Coaxum, R. A., Evans, S. C., Thomassin, K., Hersh, J., Lee, E. H., Ng, M. Y., Lau, N., Raftery-Helmer, J. N., & Mair, P. (2019): Efficient monitoring of treatment response during youth psychotherapy: The behavior and feelings survey. *Journal of Clinical Child & Adolescent Psychology*, DOI: 10.1080/15374416.2018.1547973
28. World Health Organization (2016). *Process of translation and adaptation of instruments*. World Health Organization (http://www.who.int/substance_abuse/research_tools/translation/en/, accessed 10.10.2016)
29. World Health Organization (2002). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [psychometricpropertiesofBPMandBFSappendix1.docx](#)
- [psychometricpropertiesofBPMandBFSappendix2.docx](#)