Psychotherapy Research

Psychometric Evaluation of the Behavioral Health Questionnaire-20: A Brief Instrument for Assessing Global Mental Health and the Three Phases of Psychotherapy Outcome

S.M. Kopta & J.L. Lowry

Available online: 23 Apr 2010

To cite this article: S.M. Kopta & J.L. Lowry (2002): Psychometric Evaluation of the Behavioral Health Questionnaire-20: A Brief Instrument for Assessing Global Mental Health and the Three Phases of Psychotherapy Outcome, Psychotherapy Research, 12:4, 413-426

To link to this article: http://dx.doi.org/10.1093/ptr/12.4.413

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
This study evaluated the psychometric properties of the 4 scales of the Behavioral Health Questionnaire-20 (BHQ-20): Global Mental Health, Well-Being, Symptoms, and Life Functioning. Four samples were used: community adults, college students, college students in counseling, and adults in outpatient psychotherapy. Support was found for internal consistency and 2-week test–retest reliability. For construct validity, the BHQ-20 scales distinguished levels of pathology among the samples and were sensitive to improvement across 3 psychotherapy sessions. Concurrent validity correlations with the scales of established measures were at least moderately high. Higher correlations were also found between the BHQ-20 scales and the nonanalogous scales of the established measures. These results, along with high correlations among the BHQ-20 scales, indicated the presence of 1 primary mental health dimension accounting for the variance. The BHQ-20 was proven to be a generally reliable, valid questionnaire that is distinguished by the efficiency with which it assesses mental health.

The assessment of mental health over the years has taken a variety of forms and served several purposes. The most popular method is through the self-report measure completed by the patient. This approach is more cost- and time-effective compared with other strategies such as using clinical interviews, projective techniques, and physiological measures. It is also amenable to actuarial methods of scoring and interpretation.

Mental health assessment through self-report increased in popularity beginning in the 1970s. Initially, it was used to prove that psychotherapy is generally effective, compare the effectiveness between the various therapies, and unravel the mecha-
nisms underlying therapeutic change. It was also used in mental health settings as a measure for quality assurance to meet the program evaluation requirements of community mental health construction and staffing grants. Currently, self-report assessment is serving additional roles. First, managed care’s influence has created the need to efficiently evaluate outcomes to contain costs while maintaining quality care. Care managers are relying more on self-report strategies and the resulting data to determine the allocation of resources. Second, there is a growing awareness that the integration of mental health care with primary medical care reduces the cost of and increases patient satisfaction with medical services (Lipsitt, 1997). Consequently, both physicians and psychologists are promoting the value of self-report mental health assessment in primary care settings (Maruish, 2000).

The earlier self-report instruments emphasized thoroughness. Researchers typically studied psychotherapy processes and outcomes in controlled clinical trials by giving batteries of objective tests that could take hours to complete. Popular instruments designed to give a comprehensive assessment of some aspect of mental health included the following: the 566-item Minnesota Multiphasic Personality Inventory for pathology of the personality, the 127-item Inventory of Interpersonal Problems (Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988) for interpersonal functioning, and the 90-item Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1977) for the variety of psychological symptoms. Specialized domains of interest were investigated with equal thoroughness. For example, the Beck Depression Inventory (Beck, Rush, Shaw, & Emery, 1979) used 21 items to assess depression; the 40-item State-Trait Anxiety Inventory screened for symptoms of anxiety (Spielberger, Gorsuch, & Luchene, 1970); and Wolpe and Lazarus (1966) offered a 76-item Fear Survey Schedule to evaluate fear.

In the 1990s, with the advent of managed health care and the move toward studying psychotherapy in natural settings, such as clinics and private practices, extensive testing using many items became unfeasible. Thus, briefer instruments were developed that were comprehensive in nature and assessed multiple dimensions. The most popular questionnaires include the 45-item Outcome Questionnaire-45.2 (Lambert, Hansen, et al., 1996) and the 32-item Behavior and Symptom Identification Scale (Eisen, Grob, & Klein, 1986).

Also during this time, Kenneth Howard led the development of a comprehensive psychotherapy outcomes system characterized by scales related to the process of change across sessions. Featuring a core mental health index of three scales totaling 68 items, COMPASS Treatment Assessment System (Howard, Brill, Lueger, & O’Mahoney, 1992) was designed to assess three sequentially dependent phases of improvement: well-being, symptomatic distress, and life functioning (e.g., work, family, life enjoyment). Several researchers from Dr. Howard’s group (i.e., Howard, Lueger, Maling, & Martinovich, 1993; Lutz, Lowry, Kopta, Einstein, & Howard, 2001) have empirically confirmed that greater well-being occurs first across psychotherapy sessions (remoralization), followed by symptomatic relief (remediation), and finally better life functioning (rehabilitation). The COMPASS system also included pretreatment predictive variables as well as a tracking strategy to monitor mental health over time (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Leon, Kopta, Howard, & Lutz, 1999). COMPASS is the only outcomes questionnaire that reflects, within its scales, a validated guiding theory for the process of therapeutic improvement; the scales not only indicate what is happening with the patient but, by nature of the items, when it is likely to happen. This feature allows for the systematic selection of different treatment goals and time frames depending on the desires of the parties of interest (e.g., patient, clinician, case manager).
The aforementioned measures have been demonstrated to be reliable and valid, which are two of the essential criteria for assessment. Another relevant criterion is the efficiency with which assessment occurs. For mental health questionnaires used for tracking outcomes across psychotherapy sessions or evaluating mental health across primary care appointments, brevity acquires an important role.

This study assessed the psychometric properties of a brief mental health questionnaire that measures the three phases of psychotherapy outcome as discovered by Howard et al. (1993). The following groups were assessed with the Behavioral Health Questionnaire-20 (BHQ-20): community adults, undergraduate college students, undergraduate college students in counseling, and adults in outpatient psychotherapy. The findings provided information on reliability (i.e., internal consistency, test–retest), construct validity (i.e., sensitivity to pathology, sensitivity to change), concurrent validity, and independence of the scales. We expected internal consistency and test–retest reliability coefficients to be at satisfactory levels.

For construct validity, sensitivity to pathology was assessed by using analyses of variance (ANOVAs) to compare the means of the four groups for the BHQ-20 scales: Global Mental Health, Well-Being, Symptoms, and Life Functioning. It was hypothesized that those participants in psychotherapy/counseling (i.e., college counseling clients, psychotherapy outpatients) would appear more disturbed on the scales than untreated participants (i.e., community adults, college students). Sensitivity to change was determined by assessing the amount of improvement achieved for the adult outpatient sample from intake to Session 3 of psychotherapy. It was predicted that the BHQ-20 scales would reflect positive change after three sessions of treatment.

For concurrent validity, we correlated the BHQ-20 with four established self-report measures of mental health: Behavior and Symptom Identification Scale-32, COMPASS Treatment Assessment System, Outcome Questionnaire 45.2, and SCL-90-R. We expected at least adequate correlations between the corresponding scales of these measures. Ideally, each BHQ-20 scale should correlate highest with its respective criterion measures (e.g., BHQ-20 Well-Being scale with the COMPASS Current Well-Being scale), and each scale should demonstrate low correlations with the nonanalogous measures (e.g., BHQ-20 Well-Being scale with the COMPASS Current Life Functioning scale).

Method

Participants

Participants included four adult samples: community adults not in psychotherapy, college undergraduate students not in psychotherapy, college undergraduate students in treatment at a college counseling center, and adult outpatients in psychotherapy. The community sample was composed of 380 adults, age 18 or older, not currently in psychotherapy. The sample was composed primarily of married (58.7%), Caucasian (85.0%) females (63.4%), with a mean age of 42.8 years ($SD = 13.5$). Ninety-nine percent of participants had completed high school, and 82.9% were employed full time. Slightly more than half of the participants reported that they had never received mental health treatment (54.2%).

The outpatient psychotherapy sample was obtained from a midwestern mental health center for women. The 211 patients were primarily Caucasian (78.7%) females.
(100%) with a mean age of 31.3 years ($SD = 9.1$). Ninety-four percent of patients had completed high school, 40.9% percent were employed full-time, and 16.4% were married.

Participants in the college student sample were 465 undergraduate students from a small, mid-Atlantic liberal arts college and a similar midwestern college. The college student sample was composed primarily of single (99%), Caucasian (94%) females (70%), who had not been in previous psychological treatment (87%), with a mean age of 20.4 years ($SD = 1.5$).

The students in counseling sample consisted of 208 undergraduate students who sought treatment at the college counseling center of a small, mid-Atlantic liberal arts college. The majority of student-clients were single (96%), Caucasian (88%) females (72%) with a mean age of 20.1 ($SD = 2.4$). Forty-eight percent of the sample reported previous counseling or psychotherapy experience. College counseling clients and college student participants did not significantly differ on any of the aforementioned demographic variables.

**Measures**

Behavioral Health Questionnaire-20 (BHQ-20). The BHQ-20 is a 20-item self-report measure that assesses mental health with the following scales: Well-Being (3 items), Psychological Symptoms (13 items), and Life Functioning (4 items). A Global Mental Health scale is composed of adding all 20 items. The Well-Being scale is used to evaluate emotional distress, motivation/energy, and life satisfaction. The Symptoms scale includes items that assess depression, anxiety, drug/alcohol abuse, and risk (harm to oneself and to others). Life Functioning areas are work/school, intimate relationships, nonfamily relationships, and life enjoyment. Participants are asked to rate the items regarding how they have been feeling over the past 2 weeks on a Likert-type scale ranging from 0 (extreme distress/poor functioning) to 4 (no distress/excellent functioning).

The items were selected based on the following characteristics and procedures: (a) They represent the most common problems seen in outpatient psychotherapy; (b) their face validities are consistent with the three phases of outcome (Howard et al., 1993); (c) the authors’ experience with earlier measures was used; and (d) except for the risk and drug abuse items, they had the highest level of endorsement in a large psychotherapy outpatient sample ($n = 1,417$).

Items of the BHQ-20 were administered with the following measures to establish concurrent validity: Behavior and Symptom Identification Scale-32 (Eisen et al., 1986), COMPASS Treatment Assessment System (Howard et al., 1992), Outcome Questionnaire-45.2 (Lambert, Hansen et al., 1996), and SCL-90-R (Derogatis, 1977, 1983, 1994). The aforementioned measures were selected for comparison because they assess well-being, symptomatic distress, or some form of life/interpersonal functioning.

Behavior and Symptom Identification Scale-32 (BASIS-32). The BASIS-32 (Eisen et al., 1986) is a 32-item, self-report measure designed to assess a patient’s difficulty with psychological symptoms and functioning over the past week. The instrument was empirically derived and assesses five major domains: Relation to Self/Others,
Depression/Anxiety, Daily Living/Role Functioning, Impulsive/Addictive Behavior, and Psychosis. In addition, the subscales may be averaged for a total score. Patients are to rate their level of distress on a Likert-type scale ranging from 0 (no difficulty) to 4 (extreme difficulty). Normative data are available for psychiatric inpatients and outpatients. Internal consistency data suggest that the instrument has mostly adequate reliability within the factors (Cronbach’s α = .77, .79, .76, .68, and .43, respectively, regarding Relation to Self/Others, Daily Living/Role Functioning, Depression/Anxiety, Impulsive/Addictive Behaviors, and Psychosis subscales; Eisen, Dill, & Grob, 1994). Two to 3-day test–retest reliability coefficients for a subsample of 40 patients were .80 for Relation to Self/Others, .81 for Daily Living Skills/Role Functioning, .78 for Depression/Anxiety, .65 for Impulsive/Addictive Behavior, and .76 for Psychosis. The average test–retest reliability for the instrument was .85 (Eisen et al., 1986). Concurrent validity was established by predicting patients’ return to hospitalization within 6 months based on their admission scores on the BASIS-32. Patients who reported more distress on intake were more likely to return for hospitalization than those who reported less distress. In addition, the BASIS demonstrates adequate sensitivity to change in distress across time (Eisen & Dickey, 1996; Eisen et al., 1986).

COMPASS Treatment Assessment System (COMPASS). COMPASS is marketed and distributed by Integra Inc., a managed-care company in King of Prussia, Pennsylvania. Its most recent version features several self-report scales, including (a) Therapist Rating (3 items), (b) Need for Treatment (3 items), (c) Presenting Problems (24 items), (d) Current Well-Being (CWB; 4 items), (e) Current Symptoms (CS; 33 items), and (f) Current Life Functioning (CLF; 17 items). The CWB, CS, and CLF scales combine to generate a Mental Health Index (MHI). The scales have good psychometric properties. For example, for internal consistency, the following coefficients have been reported using a sample of 423 outpatients: .79 (CWB), .60 to .88 (seven subscales of the CS), .93 (CLF), and .87 (MHI). Detailed reliability and validity information is presented by Howard et al. (1992).

Outcome Questionnaire-45.2 (OQ). The OQ (Lambert, Hansen, et al., 1996) is a 45-item self-report survey designed to assess a person's psychological and social functioning. The measure is composed of three subscales: Symptom Distress, Interpersonal Relations, and Social Role Performance. Clients are asked to complete the questionnaire with respect to how they have been feeling over the past week. Items are rated on a 5-point Likert-type scale ranging from never to almost always. Item responses may be summed to generate scores for each of the domains assessed as well as an overall score for the measure. Scores below 63, 36, 15, and 12 are indicative of healthy functioning regarding the Total, Symptom Distress, Interpersonal Relations, and Social Role Performance scales, respectively (Lambert, Okiishi, Finch, & Johnson, 1998). The OQ has the following normative data available: community adults, college students, counseling center clients, outpatients, and inpatients. Concurrent validity data showed that the OQ total score and subscales had a correlation with other similar measures ranging from .49 to .92 (Lambert et al., 1996). Test–retest reliabilities ranged from .78 to .84; internal consistency reliabilities varied from .70 to .93 for the total scale and subscales.

SCL-90-R. This is a 90-item self-report psychological symptom distress inventory (Derogatis, 1977, 1983, 1994). The measure includes three global indexes of
distress as well as nine symptom dimensions: somatization, obsessive–compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Patients rate distress experienced over the past week on a Likert-type scale ranging from 0 (not at all) to 4 (extremely). Clinically significant distress is indicated by t scores of 63 or higher on the Global Severity Index (GSI) or on a symptom scale. Normative data are available, by gender, for psychiatric outpatients, community nonpatients, psychiatric inpatients, and community adolescents (Derogatis, 1994). The SCL-90-R, one of the most widely used symptom inventories, has demonstrated good construct and concurrent validity as well as internal consistency and test–retest reliability (see Derogatis, 1994, and Derogatis & Savitz, 2000, for reviews).

Procedures

The community adult sample was obtained through two methods. One sample of participants (n = 100) was solicited from two shopping malls in a mid-Atlantic metropolitan city. Potential participants were informed that their involvement was completely voluntary and anonymous and that completion of the questionnaires, on site, was indicative of their informed consent. Each person was offered $2.00 for participation and entered into a lottery for one of four $50.00 gift certificates to be distributed at the completion of the project. The second community adult sample was solicited by a mailing to 1,004 employees of a small, mid-Atlantic liberal arts college. Each potential participant received a letter as to the purpose of the study, a demographic sheet, the BHQ-20 items, a problem checklist, and a raffle entry sheet. Participants were asked to complete the forms and return them to Jenny L. Lowry at their earliest convenience. Return of the packet was indicative of participants' informed and voluntary consent; these participants were also entered into the lottery. A total of 320 persons returned the requested information (return rate of 32%). No significant differences were found among the adult community samples; therefore, the data were merged. To ensure that the final sample would be most representative of adults not in treatment, all participants who indicated they were currently in psychotherapy were omitted from analyses (n = 40). The final sample consisted of 380 participants.

The college student sample was collected from small, midwestern (n = 327) and mid-Atlantic (n = 138) liberal arts colleges. Students were offered extra credit for a specified course in return for their participation. All students completed the items from the BHQ-20 to establish normative data. A subset of the sample (n = 184) also completed the BASIS-32, OQ-45, and SCL-90-R to establish concurrent validity of the instrument. Another subset (n = 158) completed the COMPASS instrument for the same purpose. A third subset of the sample (n = 168) completed the items of the BHQ-20 a second time, 2 weeks later, to establish test–retest reliability.

The college counseling sample (n = 206) was collected through a small, mid-Atlantic college counseling center. Clients were asked to complete the items of the BHQ-20 before their initial intake evaluation. The data were obtained before therapeutic contact to control for treatment effects.

The adult outpatient sample (n = 211) was collected at a mid-western community mental health center specializing in the treatment of females. All patients were given the items of the BHQ-20 at intake; 154 completed them at psychotherapy Session 3.
Results

Reliability

Reliability was assessed using internal consistency (coefficient alpha) analyses for each of the four samples: community adults (n = 380), college students (n = 465), college counseling clients (n = 206), and psychotherapy outpatients (n = 211). The internal consistency coefficients ranged as follows: Global Mental Health, .89 to .90; Well-Being, .65 to .74; Symptoms, .85 to .86; and Life Functioning, .72 to .77. Except for the Well-Being scale for psychotherapy outpatients (α = .65), these coefficients were at adequate levels (i.e., ≥ .72) and comparable to what has been found for similar but longer measures (e.g., BASIS-32, Eisen et al., 1994; Brief Symptom Inventory, Derogatis & Melisaratos, 1983; OQ, Lambert, Hansen, et al., 1996).

Only the college student sample was assessed for test–retest reliability. The following coefficients were significant (p < .001) and based on a 2-week interval between tests: Global Mental Health, .80; Well-Being, .71; Symptoms, .83; and Life Functioning, .80.

Construct Validity

Construct validity was assessed through the discriminant validity method of evaluating sensitivity to pathology. The statistics for this discriminant validation included a one-way ANOVA conducted on the four scales of the BHQ-20. As predicted, significant differences (p < .0001) were found among the groups for the four scales: Global Mental Health, F(3, 1258) = 179.38; Well-Being, F(3, 1257) = 154.27; Symptoms, F(3, 1257) = 141.78; Life Functioning, F(3, 1252) = 128.60. Pairwise comparisons were performed between the groups within each of the scales. The Tukey honestly significant difference (HSD) was used to control for familywise error resulting from multiple comparisons (Tukey, 1977). All pairs of means within each group were significantly different (p < .0001). Table 1 shows the group means and standard deviations for the different scales.

Assessing sensitivity to change in response to psychotherapy was another approach used to determine construct validity. Our expectations were supported by the findings. Using t-test analyses, statistically significant improvement was found in comparing intake means with Session 3 means for the four scales: Global Mental Health, M = 2.29 (SD = .70) versus M = 2.62 (SD = .81), t(153) = 5.09, p < .0001; Well-Being, M = 1.36 (SD = .86) versus M = 2.01 (SD = .97), t(149) = 6.79, p < .0001; Symptoms, M = 2.34 (SD = .84) versus M = 2.70 (SD = .90), t(149) = 5.06, p < .0001; Life Functioning, M = 1.97 (SD = .92) versus M = 2.49 (SD = .97), t(148) = 5.88, p < .0001. The effect sizes were large as defined by Lipsey (1990): Global Mental Health = .58, Well-Being = .78, Symptoms = .58, and Life Functioning = .68.

Concurrent Validity

Assessment of concurrent validity involved comparing each of the BHQ-20 scales with their respective criterion counterparts. The counterpart matches were as follows: BHQ Global Mental Health Scale with total scales of the other measures; BHQ-20 Well-Being scale with COMPASS Current Well-Being Scale; BHQ-20 Symptoms scale with BASIS-32 Anxiety/Depression, COMPASS Current Symptoms, OQ Symp-
TABLE 1. Comparisons of Level of Pathology Across Normal and Clinical Samples for the Behavioral Health Questionnaire-20 (BHQ-20) Scales

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHQ-20 Global Mental Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community adults</td>
<td>380</td>
<td>3.32</td>
<td>.47</td>
</tr>
<tr>
<td>College students</td>
<td>465</td>
<td>3.13</td>
<td>.51</td>
</tr>
<tr>
<td>College counseling clients</td>
<td>206</td>
<td>2.68</td>
<td>.62</td>
</tr>
<tr>
<td>Psychotherapy outpatients</td>
<td>211</td>
<td>2.33</td>
<td>.68</td>
</tr>
<tr>
<td>BHQ-20 Well-Being</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community adults</td>
<td>379</td>
<td>2.69</td>
<td>.76</td>
</tr>
<tr>
<td>College students</td>
<td>465</td>
<td>2.46</td>
<td>.75</td>
</tr>
<tr>
<td>College counseling clients</td>
<td>206</td>
<td>1.81</td>
<td>.85</td>
</tr>
<tr>
<td>Psychotherapy outpatients</td>
<td>211</td>
<td>1.40</td>
<td>.84</td>
</tr>
<tr>
<td>BHQ-20 Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community adults</td>
<td>380</td>
<td>3.43</td>
<td>.55</td>
</tr>
<tr>
<td>College students</td>
<td>465</td>
<td>3.22</td>
<td>.60</td>
</tr>
<tr>
<td>College counseling clients</td>
<td>206</td>
<td>2.45</td>
<td>.89</td>
</tr>
<tr>
<td>Psychotherapy outpatients</td>
<td>211</td>
<td>2.12</td>
<td>.95</td>
</tr>
<tr>
<td>BHQ-20 Life Functioning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community adults</td>
<td>379</td>
<td>3.08</td>
<td>.66</td>
</tr>
<tr>
<td>College students</td>
<td>465</td>
<td>2.85</td>
<td>.65</td>
</tr>
<tr>
<td>College counseling clients</td>
<td>203</td>
<td>2.39</td>
<td>.77</td>
</tr>
<tr>
<td>Psychotherapy outpatients</td>
<td>209</td>
<td>1.94</td>
<td>.93</td>
</tr>
</tbody>
</table>

This study assessed the psychometric properties of the BHQ-20 using samples from four distinct populations. For the most part, the findings provided support for...
our hypotheses. The instrument's scales typically demonstrated adequate to high internal consistency; test–retest reliability was good. Support for construct validity was found; all scales detected predicted differences in levels of pathology across different groups as well as improvement across three psychotherapy sessions.

Concurrent validity was good; mostly high correlations were found between the BHQ-20 scales and the counterpart scales from other established questionnaires. The BHQ-20 scales correlated highest with their respective counterparts on the other measures (i.e., Global Mental Health, Well-being, Life Functioning) or with other measures' total scores (i.e., Symptoms) with the counterparts being second highest. These results were close to what was expected. Unfortunately, correlations between the BHQ-20 scales and the nonanalogous scales of established measures were moderate to high, suggesting that the BHQ scales are not measuring distinct attributes. These findings are consistent with what other studies have reported with a variety of instruments (e.g., Hilsenroth, Ackerman, & Blagys, 2001; Umphress, Lambert, Smart, Barlow, & Clouse, 1997). Furthermore, the high correlations between the BHQ scales themselves provide additional evidence that they lack independence.

Many “multidimensional” mental health questionnaires have been developed in which the items within each scale are face valid, appearing relevant to the scale’s constructs. These measures feature an appealing theory of the multidimensionality of mental health. Typically, to their disappointment, researchers have found high correlations between the scales and concede that the scales may be essentially unidimensional, that is, tied together by a single powerful factor. The SCL-90-R (Cyr, McKenna-Foley, & Peacock, 1985), IIP (Horowitz et al., 1988), BASIS-32 (Eisen et al.,

### TABLE 2. Pearson Product–Moment Correlation Coefficients Between the Scales of the Behavioral Health Questionnaire-20 and Four Established Mental Health Instruments for the College Sample

<table>
<thead>
<tr>
<th>Scale</th>
<th>BHQ-20</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global Mental Health</td>
<td>Well-Being</td>
<td>Symptoms</td>
<td>Life Functioning</td>
</tr>
<tr>
<td>BASIS-32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>–.83</td>
<td>–.67</td>
<td>–.77</td>
<td>–.65</td>
</tr>
<tr>
<td>Anxiety/Depression</td>
<td>–.79</td>
<td>–.69</td>
<td>–.71</td>
<td>–.55</td>
</tr>
<tr>
<td>Daily Living/Role Functioning</td>
<td>–.77</td>
<td>–.61</td>
<td>–.65</td>
<td>–.65</td>
</tr>
<tr>
<td>COMPASS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>–.76</td>
<td>–.73</td>
<td>–.61</td>
<td>–.64</td>
</tr>
<tr>
<td>Current Well-Being</td>
<td>–.51</td>
<td>–.84</td>
<td>–.66</td>
<td>–.41</td>
</tr>
<tr>
<td>Current Symptoms</td>
<td>–.71</td>
<td>–.70</td>
<td>–.74</td>
<td>–.55</td>
</tr>
<tr>
<td>Current Life Functioning</td>
<td>–.71</td>
<td>–.65</td>
<td>–.55</td>
<td>–.71</td>
</tr>
<tr>
<td>OQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>–.81</td>
<td>–.62</td>
<td>–.75</td>
<td>–.65</td>
</tr>
<tr>
<td>Symptom Distress</td>
<td>–.76</td>
<td>–.56</td>
<td>–.72</td>
<td>–.60</td>
</tr>
<tr>
<td>Interpersonal Relations Plus</td>
<td>–.81</td>
<td>–.65</td>
<td>–.64</td>
<td>–.67</td>
</tr>
<tr>
<td>Social Role</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCL-90-R</td>
<td>–.85</td>
<td>–.65</td>
<td>–.85</td>
<td>–.61</td>
</tr>
</tbody>
</table>

Note. n = 184 for BASIS-32, OQ, and SCL-90-R; n = 158 for COMPASS. Correlations for counterpart scales are in italics. BHQ-20 = Behavioral Health Questionnaire-20; BASIS-32 = Behavior and Symptom Identification Scale-32; COMPASS = COMPASS Treatment Assessment System; OQ = Outcome Questionnaire-45.2; SCL-90-R = Symptom Checklist-90-Revised.
1994), and OQ (Umphress et al., 1997) are examples. These findings make sense. Usually when people experience symptomatic discomfort they are concurrently emotionally distressed and unsatisfied with life (i.e., low well-being) and function poorly. Thus, it appears that the different mental health measures share substantial variance because of a strong primary factor.

Interestingly, there is substantial evidence that highly correlated items as well as scales respond differently to psychotherapy. For example, different items within the same syndrome (e.g., crying vs. insomnia for depression) for the SCL-90-R (Kopta, Howard, Lowry, & Beutler, 1994) and the BDI (Barkham et al., 1996) demonstrated different rates of improvement across psychotherapy sessions. COMPASS scales have shown higher intercorrelations but responded at different rates to treatment (Howard et al., 1992). Thus, the BHQ-20 items and scales show promise in reflecting the multidimensional recovery process in psychotherapy; therefore, the BHQ-20 may be clinically useful beyond being a general measure of mental health.

As has been found with similar brief scales, internal consistency reliabilities for the Well-Being and Life Functioning scales were marginally adequate because of the small number of items (three and four, respectively). In this case, the researcher's particular needs become important. The brevity of shorter measures with their lower reliability has to be weighed against the higher reliability typical of longer measures. Still, these BHQ-20 scales showed comparable reliability to their longer counterparts. For example, the 11-item Interpersonal Relations and 9-item Social Role scales of the OQ were reported to have internal consistency reliabilities ranging from .70 to .74 (Lambert, Hansen, et al., 1996). For our college student sample, the reliability coefficients were .72 for Interpersonal Relations and .63 for Social Role. The reliability of COMPASS's four-item Current Well-Being scale was slightly higher at .79 (Howard et al., 1992) and at .73 as found in the college student sample.

The findings showed support for the construct validity of the BHQ-20 with regard to the phase model. The Well-Being, Symptoms, and Life Functioning scales demonstrated the highest correlations with their COMPASS counterparts. The COMPASS scales have been the primary scales used to distinguish the different phases of change across psychotherapy sessions (Howard et al., 1993; Lutz et al., 2001). Because change was only assessed over the first three psychotherapy sessions, confirmation of the phase model in this study is limited. The phase model proposes that the expected direction of therapeutic movement across sessions is improved well-being followed by symptomatic stress reduction and then better life functioning. Thus, as expected, the largest effect size across the first three sessions was found for the Well-Being scale (.78) with less change characterizing the Symptoms (.58) and Life Functioning (.68) scales. A similar pattern, consistent with the phase theory, was reported for the COMPASS scales within the first three sessions (Howard et al., 1993; Lutz et al., 2001).

By investigating levels of psychopathology across distinct populations, we found interesting contrasts among our samples. Normal community adults looked the healthiest, including being healthier than normal college undergraduates. College counseling clients, although sicker than the two untreated groups, were not as disturbed as the psychotherapy outpatients. Apparently, these individuals remained healthy enough to struggle along socially and academically in college. Adult outpatients, who typically seek help for longer term, deeper problems of symptomatology and life functioning, were the most disturbed. Other researchers have found a similar pattern of psychopathology. For example, Umphress et al. (1997), using the OQ, reported that community clinic outpatients had greater symptom distress and interpersonal prob-
lems than students in college counseling; community normals scored at healthier levels than both of the clinical samples. However, different from our results, Lambert, Burlingame, et al. (1996) found that normal college undergraduates were healthier than normal community adults in overall levels of psychopathology as measured by the OQ.

Future research with the BHQ-20 should include determining whether the scales can detect differential responsiveness to psychotherapy beyond three sessions. Based on the phase model, it is expected that symptomatic distress, followed by life functioning, would show greater improvement than well-being as patients continue with more treatment sessions. Although it was found that the BHQ-20 scales measure sensitivity to psychotherapeutic improvement over sessions, further support for construct validity could be investigated by using a strategy recommended by Vermeersch, Lambert, and Burlingame (2000). This approach involves comparing a treatment group with a control group over time to assess whether improvement from treatment is greater than the change detected in untreated individuals. Another consideration is to compare BHQ-20 dose–effect relations with those reported by investigators using other instruments such as the SCL-90-R (Kopta et al., 1994), OQ (Kadera, Lambert, & Andrews, 1996), and BDI (Barkham et al., 1996). Here, as in the other studies, improvement could be operationally defined in terms of clinical significance and reliable change cutoff points (Jacobson & Truax, 1991).

Expected treatment response (ETR) methodologies are used to predict an expected path of patient progress across sessions before psychotherapy begins (Leon et al., 1999; Lambert, Hansen, & Finch, 2001). These methods, which are becoming increasingly popular, show promise with regard to improving the effects of psychotherapy. Sufficient between-patient and temporal variance is essential to produce the recovery curves adequately. As number of items in the measure influences this variance, the effect of the briefer BHQ-20 compared with the longer instruments typically used in ETR (e.g., COMPASS, OQ) needs to be assessed to ensure that the BHQ-20 is useful for ETR research.

This study demonstrated that the BHQ-20 is a valid and reliable brief measure that assesses symptoms common to psychotherapy outpatients as well as well-being and critical areas of life functioning (i.e., work, relationships, enjoyment). The BHQ-20 was developed with the previous work of Kenneth Howard in mind. He was aware of the project as it proceeded and provided valuable advice. The goal was to create a brief instrument to assess phase model attributes at a time when efficiency in assessment is becoming as important as comprehensiveness.

References


Zusammenfassung

Résumé
Cette étude évalue les propriétés psychométriques des 4 échelles du Behavioral Health Questionnaire (BHQ-20) : Santé Mentale Globale, Bien-Etre, Symptômes et Fonctionnement dans la Vie. Quatre échantillons ont été utilisés : citoyens adultes, étudiants, étudiants en consultation, et adultes en traitement ambulatoire. La consistance interne et la fidélité test-retest à 2 semaines se sont vues corroborées. En ce qui concerne la construct validity, les échelles du BHQ-20 ont distingué les niveaux de pathologie entre les échantillons et ont été sensibles à l'amélioration à travers 3 séances de psychothérapie. Les corrélations de la validité concurrente avec des échelles de mesures établies étaient au moins modérément élevées. Des corrélations plus hautes ont été trouvées également entre les échelles BHQ-20 et les échelles non analogiques de mesures établies. Ces résultats plus des corrélations hautes entre les échelles BHQ-20 indiquent la présence de 1 dimension de santé mentale primaire qui explique la variance. Le BHQ-20 s’est avéré un questionnaire dans l’ensemble fidèle et valide qui se démarque par l’efficience avec laquelle il évalue la santé mentale.

Resumen
Este estudio evaluó las propiedades psicométricas de las cuatro escalas del Cuestionario de Salud Conductal-20 (BHQ-20): Salud mental global, Bienestar, Síntomas y Desempeño en la vida. Se utilizaron cuatro muestras: adultos de la comunidad, estudiantes universitarios de pregrado, estudiantes de counseling y adultos ambulatorios en psicoterapia. Se encontró consistencia interna y confiabilidad en un “test-retest” a las dos semanas. En cuanto a la validez del constructo, las escalas del BHQ-20 diferenciaron niveles de patología entre las muestras y acusaron mejoría en tres sesiones de psicoterapia. Las correlaciones de validez concurrente con las escalas de medidas establecidas fueron moderadamente altas. Se encontraron correlaciones más altas entre las escalas BHQ-20 y las escalas no análogas de las medidas establecidas. Estos resultados, junto con altas correlaciones entre las escalas BHQ-20, indicaron que la varianza se explica por la presencia de una dimensión primaria de salud mental. La BHQ-20 probó ser un cuestionario válido y confiable en general, que se distingue por la eficiencia con la cual puede evaluar la salud mental.

Resumo
Este estudo avaliou as propriedades psicométricas das 4 escalas do Questionário de Saúde Comportamental-20 (BHQ-20): Saúde mental geral; Bem-estar, Sintomas e Funcionamento. Foram utilizadas quatro amostras: adultos na comunidade, estudantes universitários, estudantes universitários em consulta, e adultos em psicoterapia. Foi encontrado apoio para a consistência interna e fidelidade teste-reteste de duas semanas. Para a validade do constructo, as escalas do BHQ-20 distinguiram níveis de psicopatologia entre as amostras e foram sensíveis à melhoria ao longo de três sessões de psicoterapia. Correlações de validade concurrente com escalas de medidas estabelecidas foi pelo menos moderadamente elevada. Correlações mais elevadas foram encontradas entre as escalas do BHQ-20 e as escalas não análogas das referidas medidas estabelecidas. Estes resultados, em conjunto com as elevadas correlações entre as escalas do BHQ-20, indicam a presença de uma dimensão primária de saúde mental.
que explica a variância. O BHQ-20 provou ser, globalmente, um questionário fiável, válido que se
ningue pela eficiência com que avalia a saúde mental.

Sommario
Lo studio ha valutato le proprietà psicometriche delle 4 scale del Questionario Attitudinale-20 (BHQ-20):
Salute Mentale Globale, Beneessere, Sintomi, e Funzioni Vitali. Sono stati utilizzati quattro campioni: adulti
di comunità, studenti di college, studenti di college in terapia, ed adulti in psicoterapia ambulatoriale. È
stata riscontrata una coerenza interna ed un’affidabilità per 2 settimane dal test-retest. Per la validità del
costutto, le scale del BHQ-20 distinguovevano i campioni in vari livelli di patologia ed erano sensibili al
miglioramento nell’arco di 3 sedute di psicoterapia. Correlazioni di validità concomitante con le scale delle
misure stabilite sono risultate almeno moderatamente elevate. Delle correlazioni più elevate sono state
riscontrate tra le scale BHQ-20 e le scale non analoghe delle misure stabilite. Questi risultati, assieme alle
elevate correlazioni tra le scale BHQ-20, hanno indicato la presenza di 1 dimensione mentale primaria
che giustificava la varianza. Il BHQ-20 si è dimostrato un questionario generalmente affidabile e valido
che si contraddistingue grazie all’efficacia con cui valuta la salute mentale.

Received July 31, 2001
Revision received June 5, 2002
Accepted June 5, 2002