CREATING A BIO-PSYCHO-SOCIAL-SPIRITUAL MEASURE OF PHYSICIAN WELL-BEING

Introduction

This is a summary of the development of the Physician Well-Being Self-Assessment Tool (PWSAT) (Copyright © AHSS dba Florida Hospital, 2010, available exclusively from Courageous Healthcare, Inc.). This summary quickly reviews the development process and early results on reliability and validity as of April 2016. Development of the PWSAT was completed with the support of Ted Hamilton, MD, Senior Vice President of Medical Mission and underwritten by Adventist Health System-Sunbelt, which does business as Florida Hospital. Richard J. Bogue, PhD, FACHE, led a research team in support of a national expert panel of physicians, psychologists and other researchers.

Development of the PWSAT took place in the chronology typical of scale construction. First, a purpose for scale construction was conceived and clarified. Second, the resulting concept was examined by drawing on existing evidence to ensure content and face validity. Third, a formal literature review ensured broad and deep coverage of the sources of information for potential scale items. Fourth, systematic and independent ratings of items by an Expert Panel reduced the number of candidate items and identified subscale membership for the remaining items. After these initial validity steps, the scale was subjected to initial pilots to further reduce the number of items and to improve the reliability and efficiency of the scale; that is, to maximize the fit between the concept and the scale and to do so with the smallest set of items for the scale's purpose. Finally further testing was conducted, and is conducted on an ongoing basis, to continue evaluating internal reliability and to establish external validation by evaluating associations with other measures.

Content and Face Validity

A survey has ***content validity*** when systematic procedures are used to ensure that the survey is developed based on the right information. To ensure ***content validity*** for a measure of physician well-being, the research team conducted a comprehensive literature review on factors affecting physician well-being. The literature search strategy sought studies reporting empirical results about impacts on some aspect of "health" and/or "well-being" for physicians in particular. Only peer-reviewed articles and research-based books from 1990 through 2007 were initially identified (N = 115). Articles cited by these sources were reviewed to identify reports on methods of measuring physician health and/or well-being.

Face validity means that content experts and users judge the survey as being about what it is purported to be about. Face validity for PWSAT was assured systematically and iteratively. Factors reported to have an impact on physician health or well-being were extracted from the articles, yielding 261 candidate items. These were supplemented by 96 items from the Behavioral Risk Factor Surveillance Survey (CDC), producing a total of 357 candidate items. Four research team members reviewed these 357 candidate items to eliminate (a) duplicates, (b) near duplicates, (c) items that were not structurally neutral as to gender or race (eg, pregnancy or prostate cancer), (d) items that could not

be used in a Likert-type scale, and (e) items lacking face validity as indicators of good or poor well-being.

After this procedure, 122 items remained on the candidate list. These 122 items were reviewed by a five-member expert panel to gain a sense of the underlying concepts within well-being. The panel developed a working definition of wellbeing, the current version of which reads: *Physician well-being is a sense of satisfaction and fulfillment that characterizes personal and professional life. Well-being embraces concepts of health, contentment, balance, meaning and purpose. It is evidenced by lifestyle and behaviors that contribute to personal wholeness and healing. These practices include caring for oneself and one's family, nurturing professional relationships with colleagues and other caregivers, and treating everyone with respect, kindness, appreciation, and empathy.* The panel also identified and affirmed four possible domains of well-being: bio-physical, psycho-emotional, socio-relational, and religio-spiritual wellbeing.

Conceptual Structure and Focus

For more formal assessments to validate the conceptual structure of PWSAT, the expert panel was expanded to include five additional physicians with leadership roles in medical education in four different health systems. A third PhD researcher with extensive experience in cultural and spiritual aspects of patient care was also added. Through an online survey, the expanded expert panel members (N = 11) independently rated the 122 candidate items as follows: (a) To which of the four possible domains does the candidate item belong? (b) How important to health is the candidate item? (c) How relevant to physicians in particular is the item? (d) Is the item of an appropriate level of specificity to be used in a self-report survey instrument? (e) To what degree can one discern whether the item's impact would be helpful or harmful to health?

Items were *allocated to domains* of wellbeing based on the expert panel's ratings. Within each domain of well-being, items above the mean on *importance to health* were retained. This produced a minimum of 15 candidate items in each domain, except the bio-physical; in this domain, additional items were retained if they either exceeded the mean rating for *physician relevance* or if at least 8 of the 11 reviewers independently indicated that the item had "about the right *specificity*." This process retained 87 total items, roughly balanced across the four domains. Any items that were originally unclear as to whether its expression would be helpful or harmful to health were evaluated and reworded as needed to ensure clarity.

Internal Reliability

After a survey instrument is determined to have content and face validity, it can begin to be tested by gathering data with it and evaluating the instrument based on those data. First, several small pilots had medical students complete the 87-item instrument to assess and finalize item and scale wording and to test different scaling options. The 87 item-instrument was then tested with a convenience sample of 120 physicians planning to attend a conference on physician well-being. Participants rated "How accurately each statement describes you" on a 9-point semantic differential scale from Extremely

Accurate (9) to Extremely Inaccurate (1). While only minor questions arose on this score, after this pilot, these wordings were modified to improve clarity for respondents to become, as in the version used since: Extremely True of Me (9) to Extremely Untrue of Me (1).

In this pilot, item order was randomized for each participant. Factor analysis forcing a fit to four factors with principal axis extraction and equamax rotation verified that items loaded primarily in the intended groupings or domains. Reliability analyses were used to remove items that were not contributing to the relevant domain scale's Cronbach's Alpha reliability measure. This process of eliminating poorest fitting items was continued until resolving on the best-fitting 10 items for each domain.

Table 1 below summarizes ***reliability*** results for the first three studies using PWSAT. For each study, the average item-scale correlation is shown (Avg r) as well as the Cronbach's Alpha measure of scale reliability. For Study3, Guttman's Split-Half procedure was also used. These early results demonstrate generally strong reliability.

The Study3 results may provide early evidence of modest opportunities to improve the performance of the BIO and RELA scales in particular. However, it should be pointed out that Study3 should have been expected to produce somewhat lower reliabilities as respondents were randomly selected from a national database of active physicians. Moreover, given the value of practicality of balanced 10-item scales for each domain, adding items is not desirable unless more evidence points to the need to do so.

Table 1: PWSAT-40, Reliability.

	Study1		Study2		Study3		
	Avg r	α	Avg r	α	Avg r	α	G
BIO	.640	.840	.592	.807	.352	.678	.675
EMO	.695	.883	.662	.859	.546	.846	.837
RELA	.577	.795	.572	.762	.367	.700	.501
SPIR	.652	.843	.569	.771	.582	.873	.908

PWSAT-40 is the long form version with 40 items, 10 for each of the four scales: Bio-Physical (BIO), Psycho-Emotional (EMO), Socio-Relational (RELA) and Religio-Spiritual (SPIR).
Study1: Physician attending the 2012 National Physician Well-Being Conference participated online prior to the conference (N = 81; 2012).
Study2: The 1,200 members of a religiously-affiliated health system medical staff were invited to participate online prior to a conference (N = 150; 2012).
Study3: National Study on Physician Well-Being randomly sampled 1,615 physicians nationwide (N = 225; 2012).
Avg r = Average item-scale correlation. All items have r > .30 on their respective subscale.
α = Cronbach's Alpha measures internal consistency ($\alpha \geq .70$ is considered adequate; $\alpha \geq .80$ good; and $\alpha \geq .90$ excellent).
G = Guttman Split Half assesses the correlation of half the items with the other half.

Initial Evidence on External Validity

Hamilton & Bogue (2012a) found a strong and linear inverse relationship between overall PWSAT scores and the Depersonalization Scale of the Maslach Burnout Inventory for Health Professionals (MBI-HP). MBI-HP has been extensively used and is widely taken as the standard measure of burnout. It has three subscales: Depersonalization (of clients/patients), Emotional Exhaustion (of the professional) and Personal Accomplishment (understood as a protective factor that mitigates against burnout among health professionals).

Depersonalization (DP) is evidenced when clinicians depersonalize their patients, for example by thinking of them or communicating with them in dismissive or disrespectful ways. DP scale scores produce diagnostic categories of Low, Medium and High depersonalization. People with High DP scores are at high risk of burnout. Among the first 143 respondents to the National Study on Physician Wellbeing, the average overall PWSAT score for physicians with Low DP was 6.16, for the Medium DP it was 5.52 and for the High DP group it was 5.11. Analysis of Variance was used to reveal this significant inverse relationship between Well-Being and DP among the first 143 respondents to the National Study ($p < 0.001$). ***Higher overall well-being as measured by PWSAT is strongly associated with lower levels of depersonalization***.

The Emotional Exhaustion (EE) component of the MBI taps the health professional's disabling emotional fatigue as the second major component of burnout. Like the DP scale, the EE scale scores produce diagnostic categories of Low, Medium and High EE, indicating Low, Medium and High burnout risk levels. Among the first 143 respondents to the National Study, 46.6% were at High risk of EE, while 26.0% were at Medium risk and 27.4% were at Low risk. Similar to the above inverse relationship between DP and PWSAT, EE and the EMO scale had a strongly significant inverse relationship ($p < 0.001$). Those in the Low EE category had EMO scores of 6.32, the Medium EE's had EMO scores of 6.00 and those in the High risk EE group had EMO scores of only 5.12. ***Higher psycho-emotional well-being as measured by PWSAT is strongly associated with lower levels of emotional exhaustion***.

Also using the first 143 respondents to the National Study on Physician Well-Being, Hamilton and Bogue (2012a) examined the convergent and discriminant validity of the four scales relative to a series of items about religious or spiritual concepts (Table 2). ***Convergent validity*** is shown when a measure strongly and positively associates with other concepts that it would be expected to associate with. As might be expected, the PWSAT Religio-Spiritual (SPIR) scale scores were significantly higher for those who see medicine as calling, those who attend services more than once per month, those who indicated that religion is important in their lives, and those who consider themselves spiritual. This finding provides evidence of ***convergent validity*** for the SPIR scale with external measures that can be expected to be associated with those for whom religion and/or spirituality is more important or central in their life.

Table 2: Evidence of Convergent and Discriminant Validity of PWSAT SPIR scale.

Measures Related to Religiosity or Spirituality	BIO	EMO	RELA	SPIR
See Medicine as a Calling	NS	NS	NS	YES
Attend Services More Once per Month	NS	NS	NS	YES
Religion is Important in My Life	NS	NS	NS	YES
Consider Myself Spiritual	NS	NS	NS	YES

It is important that measurement scales discriminate between concepts and groups of people. If a measure associates with "everything" it means it is really measuring nothing in particular. Strong evidence of ***discriminant validity*** is also shown in Table 2 above. None of the other three PWSAT scales were significantly related to any of the four external indicators of religiosity and/or spirituality.

Concurrent validity demonstrates that an instrument is able to distinguish between different groups of respondents. Initial evidence on the concurrent validity of PWSAT was gained by comparing results from the first 143 respondents for the National Study on Physician Well-Being to the results from the respondents of the medical staff of a religiously affiliated health system in Midwest. The two samples had significantly different average scores on the EMO, RELA, and SPIR scales ($p < 0.001$).

Summary (January 24, 2017)

PWSAT produces four subscale scores measuring bio-physical (BIO), psycho-emotional (EMO), socio-relational (RELA) and religio-spiritual (SPIR) wellbeing. These efficient 10-item subscales have been found to measure what they are intended to measure and to be of good to excellent internal reliability. Early evidence on external validity is promising and suggests that PWSAT can be very useful as an index of wellbeing. Additional assessments of the external validity of the PWSAT are currently underway. PWSAT has been modified and used with wording changes to create additional version for medical students (PWSAT-S), nurses (NWSAT), care givers in general (WSAT), and non-provider populations (WSAT-G). Each version of the Wellbeing Self-Assessment has undergone or is undergoing further evaluation.

References

Bogue RJ and Slockett D. (2013). "How Do You Excel at Being Well?" Adventist Health System, 2013 National Physician Well-Being Conference, Ponte Vedra, FL, April 5.

Hamilton T and Bogue RJ. (2012a). "Physician Well-Being: What We've Learned." Adventist Health System, 2012 National Physician Well-Being Conference, Ponte Vedra, FL, April 14.

Hamilton T and Bogue RJ (2012b). "Surviving Success: Wellness and the Physicians." St. Vincent Health System, 2012 Fall Physician Engagement Retreat, Indianapolis, IN, November 8.

Leiter, M. P., & Maslach, C. (2005). A mediation model of job burnout. In Antoniou, A. S., & Cooper, C. L. (Eds.), Research companion to organizational health psychology (544-564). Elgar Publishing.

Maslach, C. (2003). Job burnout: New directions in research and intervention. Current Directions in Psychological Science, 12, 189-192.

Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. Journal of Applied Psychology, 93, 498-512.

Maslach, C., Leiter, M. P., & Schaufeli, W. B. (2009). Measuring burnout. In C. L. Cooper & S. Cartwright (Eds.), The Oxford handbook of organizational well-being (86-108). Oxford UK: Oxford University Press.