

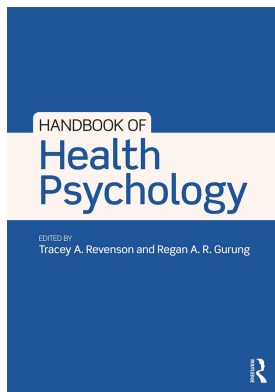
This article was downloaded by: 10.3.98.104

On: 19 Sep 2020

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Health Psychology

Tracey A. Revenson, Regan A. R. Gurung

Measurement in Health Psychology Research

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315167534-3>

Aleksandra Luszczynska, Magdalena Kruk, Monika Boberska

Published online on: 18 Dec 2018

How to cite :- Aleksandra Luszczynska, Magdalena Kruk, Monika Boberska. 18 Dec 2018, *Measurement in Health Psychology Research from: Handbook of Health Psychology* Routledge
Accessed on: 19 Sep 2020

<https://www.routledgehandbooks.com/doi/10.4324/9781315167534-3>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or 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 publisher shall not be liable for an 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.

3

MEASUREMENT IN HEALTH PSYCHOLOGY RESEARCH

*Aleksandra Luszczynska, Magdalena Kruk,
and Monika Boberska*

The process of measurement is the cornerstone of evidence-based research and practice. The strategies used to select the measures, the way the measures are implemented, and the evaluations of instruments determine *what* and *how* a researcher or practitioner can diagnose, explain, or predict.

Measurement, assessment, and evaluation are closely related terms. *Measurement* focuses on describing the characteristics of an individual (Apple, 2005, 2006). For example, health psychologists may measure the average level of adherence to medication in patients with multiple sclerosis or the strength of self-efficacy beliefs in smokers who intend to quit. Assessment may deal with a process of obtaining information in relation to internally defined goals or criteria, identifying areas of improvement or adjustments needed (Apple, 2005, 2006). Assessment is process-oriented, and uses flexible, modifiable criteria. For example, health psychologists may assess a reduction of smoking and the progress in reaching an individual's smoking-related goals.

Evaluation accounts for the individual in a specific situation and the goals or criteria that were externally set; evaluation results in a normative judgment and it is not process-oriented (Apple, 2005, 2006). Evaluation aims at a decision regarding whether a certain objective was reached. For example, health psychologists may evaluate if an intervention resulted in a clinically significant reduction of pain or whether an individual meets the physical activity recommendations formulated by World Health Organization. Differentiating between assessment, measurement, and evaluation is grounded in learning and education models (Apple, 2005). In health psychology these terms are often used interchangeably. However, approaches differentiating measurement, assessment, and evaluation are signaling the complexity of the development and use of measurement instruments.

As health psychology attempts to account for the complexities behind any act of measurement, choosing a measure requires careful planning. This chapter presents five steps that may be made when planning, selecting, and implementing the optimal instruments to measure constructs and processes viable for health psychology. These steps account for (1) the choice of a general framework, considering purpose and domains of measurement, (2) characteristics of the target population, (3) the type of measurement, (4) psychometric characteristics, and (5) issues of implementation of an instrument.

Box 3.1

Five steps in planning, selecting, and implementing a measurement instrument.

1. Choose a General Framework Guiding Measurement in Health Psychology
2. Consider the Specificity of the Target Population
3. Make a Decision About the Type of the Measurement
4. Consider the Psychometric Properties of the Measures
5. Consider Implementation of Measurement

Step 1. Choose a General Framework Guiding Measurement in Health Psychology

There are several general frameworks developed to help to plan the process of measurement. These frameworks refer to the purpose of the measurement or the domains measured (Karademas, Benyamini, & Johnston, 2016). These frameworks are, in turn, guided by a general assumption that health psychology is a discipline focusing on understanding within-individual processes that affect individual's health.

The frameworks referring to the *purpose of measurement* (e.g., Karademas et al., 2016) distinguish the following approaches to measurement:

1. Measurement conducted to reach a clinical decision about an individual. Examples may include using a measure to determine if an individual requires an intervention or if he/she has improved in a specific area of functioning as a result of an intervention.
2. Measurement aiming at describing a target population. The examples would include describing the sociodemographic characteristics, affect, level of psychosocial resources, or health status of a target patient population of people diagnosed with multiple sclerosis.
3. Measurement conducted to elucidate health-related processes that predict psychosocial outcomes. This type of assessment is focused on tapping the mechanisms responsible for future health outcomes. Usually this approach is driven by a specific theory and incorporates the constructs and processes in the theory. This approach is used to discover if empirical evidence provides support for a theory, including the assumed role of specific psychosocial predictors, the relationships between them, and their assumed links with health outcomes. The type of health outcome would be also determined by a theory. For example, a theory may focus on emotional and illness-related outcomes (the common sense model; Leventhal, Phillips, & Burns, 2016) or health behavior change (social cognitive theory; Luszczynska & Schwarzer, 2015).

Overall, purpose-related frameworks provide a general direction for selection and evaluation of an instrument, for example in terms of its efficacy, usefulness, or feasibility. Any organizing principles depend on the purpose of the measurement. For example, measurement aimed at population description should be informed by an analysis of the characteristics and needs of the target population. In contrast, when assessing psychosocial processes and their health outcomes, the organizing principle should refer to the theory-assumed proximal and distal predictors, as well as the ways they are linked together, forming mechanism responsible for changes in respective outcomes. Finally, measurement to make a diagnosis should be organized by a purpose of diagnosis within context (individual characteristics, setting, and the healthcare system).

Other general frameworks organize measurement by domains in health psychology. This approach corresponds with the main tasks of health psychology and areas of research and practice. For example, Smith (2003, 2011) proposes that measurement may be organized into three areas, described next.

1. *Domain of health behavior and prevention.* This domain includes measurement of variables from models specifying behavior change mechanisms, assessment of behavior change interventions (their processes and outcomes), the actual health behaviors (e.g., smoking), interventions aiming at behavior change, and their effects on other health outcomes (e.g., cardiovascular morbidity).
2. *Domain of stress and health.* This incorporates measurement of psychosocial variables responsible for developing illness (e.g., personality, social influence, affect, stress), and assessment of interventions aiming at reduction of the impact of these variables on stress and health.
3. *Domain of illness and care,* accounting for the measurement of adaptation to illness, the impact of illness on quality of life and well-being, as well as the evaluation of interventions aiming at patients' adjustment to illness.

Domain-specific measurement frameworks may use broad areas of health issues as the organizing principle (Smith, 2011). Selecting the area would help researchers and practitioners to focus their measurement processes on theoretical models, mechanisms, and outcomes that are established and empirically validated. This, in turn, would guide the selection of measures that fit the respective context.

In contrast to global frameworks, other approaches organize measurement in health psychology into narrow, specific domains. For example, Karademas et al. (2016) listed 16 domains of measurement, including social cognitions in health behavior, self-efficacy and outcome expectancies, health behavior, illness representation, stress and stressors, coping, social support, neuropsychological assessment, self-rated health, biological and physiological measurement, patient-physician communication and satisfaction, adherence to medical advice, pain and pain behavior, functional status outcomes, psychological adjustment, and quality of life. Such specific approaches have been developed in recognition of the interrelations or overlaps between the domains of health psychology research and practice. For example, when measuring smoking in response to stress, the global framework proposed by Smith (2003, 2011) may be less useful than Karademas's et al. (2016) approach, proposing measurement domains related to stress and health behavior.

Both purpose and domain frameworks (Karademas et al., 2016; Smith, 2003, 2011) highlight the role of the use of psychological theory the process of selection of the optimal instrument and measurement. The applied instruments should capture the processes and constructs described by the selected theory. Psychological theory would provide a specific guide about the content of the specific construct which needs to be covered; it may also specify how global (or outcome-specific) the assessment should be, the preferable response format, and in some cases, what should be the order in which the instruments are applied (Conner & Norman, 2015). Theories that provide very detailed operationalization guidelines increase the comparability between studies (Conner & Norman, 2015).

For example, when using social cognitive theory to explain a health outcome, a self-efficacy instrument should account for individual beliefs in their capability to exercise control over challenging demands and be always outcome-specific (Luszczynska & Schwarzer, 2015). The challenging demands should be population-specific and obtained in an elicitation study, conducted in a sample from the target population (e.g., adolescents). Ideally, the items measuring self-efficacy should include such phrases as "I believe that am able to," "be/perform an action such as . . .," "even if . . . (barriers/challenges occur)", for example, "I am able to adhere to my post-transplant medication even if I find it difficult to integrate it into my busy daily schedule" (Luszczynska & Schwarzer, 2015).

The two types of frameworks (purpose-oriented and domain-oriented) suggest that when selecting an instrument, health psychologists should consider multiple issues, such as the age of the

respondent, the psychometric properties reliability of the instrument, and the effect of the instrument itself on the respondents' behaviors (cf. Johnston, Benyamini, & Karademas, 2016; Smith, 2011). These issues may be organized along several dimensions: the characteristics of the target population (e.g., age, gender, ethnicity, culture), the type and format of the measurement (e.g., self-report vs objective indicators), and the psychometric properties of the measure (e.g., validity, sensitivity, factorial structure).

Step 2. Consider the Specificity of the Target Population

Age of the Target Population

Age of the target population is among the prime issues accounted for when planning the measurement.

Children and Adolescents

When the target group consists of children, the information referring to health and its psychosocial indicators may be measured through self-report by either children (aged 8–11) or adolescents (12–18 years old) only, by obtaining reports of others (parents, healthcare professionals), or both. Proxy measurement must be considered with special care, as parents may tend to overestimate illness-related psychosocial distress in their children (Vetter, Bridgewater, & McGwin, 2012).

In case of younger children or when there are concerns regarding the child's level of understanding, formal assessment of cognitive skills may be required (Christie, 2015). This may be particularly relevant when the aim of measurement is the clinical diagnosis (Christie, 2015). Several self-report instruments have been validated for children as young as 6 years old (Christie, 2015). Unfortunately, when considering instruments with established psychometric properties, the vast majority were developed to measure outcomes, that is, emotional health, quality of life, anxiety, fatigue, and chronic pain (Christie, 2015).

Older Adults

Measurement of health-related predictors and outcomes in older adults may also pose age-specific challenges. For example, an overestimation of distress severity may occur if instruments contain items asking about somatic symptoms that are likely to reflect physical comorbidities (Karp, Rudy, & Weiner, 2008). Performance of instruments assessing emotional health in older adults may be also compromised by sensory deficits, cognitive impairment, and fatigue (cf. Christie, 2015; Karp et al., 2008).

Cultural Anchoring

A second key issue refers to the role of culture, both cultural adaptations of measures as well as cultural competences of instruments. Carefully done cultural adaptation of an instrument involves semantic, idiomatic, experiential, and conceptual equivalence (Beaton, Bombardier, Guillemin, & Ferraz, 2000).

Semantic equivalence ensures that across different language versions of a measure, the meaning of the words remains the same and that grammar represents the overall characteristics of grammar in a respective language. For example, the use of passive voice might be proportionally adjusted to the use of passive voice in a respective language. *Idiomatic equivalence* refers to identifying idioms (e.g., "hard as nails" in English) and substituting them with structures parallel in the other languages. Furthermore, items are usually seeking to capture the everyday experience of daily life. In some cultures, a given situation, task, or behavior may not be experienced, in which case one item needs to be replaced by

another item capturing a situation/task that is experienced and is of a similar nature to the original task. This way, experiential equivalence is reached. Finally, *conceptual equivalence* means that the same construct is covered by the different language versions (e.g., Beaton et al., 2000). Across languages, even everyday words hold different conceptual meanings; for example, the concept of family may be different across the cultures, varying between nuclear and extended family as a prototype model.

There are two main approaches to cultural issues in measurement. The first focuses on cross-cultural differences (Betancourt, Green, Carrillo, & Firempong, 2003). According to this approach, the measure should capture culture-specific language, culture-specific constructs and processes. Research and practice sensitive to culture and ethnic diversity in the target population may require qualitative studies to precede development of a measure in order to adjust the cultural sensitivity and specificity of the instrument. In addition to identifying culturally indigenous expressions reflecting specific construct, early work would usually search for culture/ethnic group-specific aspects of constructs. For example, when attempting to measure beliefs affecting cancer screening, the qualitative study may elicit culture-specific barriers referring to fatalism (Abraído-Lanza, Cunha Martins, & Shelton, 2015). Critics of this approach argue that the large array of cultural factors make this approach inefficient and contribute to ethnic stereotyping (Betancourt et al., 2003). Furthermore, within-group variations in ethnic minority groups, in particular, socioeconomic and acculturation differences, may have stronger effects than culture or ethnic-group status (Betancourt et al., 2003).

The second approach emphasizes similarities and cross-cutting cultural issues that are present across cultures and relevant to multiple ethnic groups (Betancourt et al., 2003). Cross-cutting issues include the role of family in health and illness, racism and prejudice against minorities, decision-making and communication preferences, group-based medical mistrust, and group-level perceptions of susceptibility to cancer (Betancourt et al., 2003; Purnell et al., 2010). For example, in the context of measuring cancer treatment and prevention the instrument could account for treatment/care related mistrust, e.g., “health specialists treat people of my ethnic group like guinea pigs”.

Another term that has been coined in the context of the quality of healthcare services is *cultural competence* (Betancourt et al., 2003). Culturally competent health services would apply culturally competent instruments. Culturally competent approaches acknowledge and incorporate the importance of culture (e.g., accounting for culturally sensitive content), continuously assess effects of culture on instruments applied, and account for possible cultural differences in the interrelations between beliefs, behaviors, illness, and treatment (Betancourt et al., 2003). Importantly, any culturally competent approach differentiates cultural factors from socioeconomic factors, such as education.

Step 3. Make a Decision About the Type of the Measurement

There are many ways to measure the same construct. A decision is based on the other factors described above, such as specific characteristics of population, but also includes resources, cost, and patient burden. The most frequently applied types of assessment in health psychology will be discussed: self-report (including ecological momentary assessment), biomarker, and sensor-based instruments.

Self-Report Measures

Self-report is considered the most practical type of measurement in health psychology research, in particular in the case of large populations-based studies. It is often the method of choice for describing characteristics of a target population (“What is your highest level of education?”) or elucidating processes and mechanisms explaining health outcomes (e.g., “Please indicate how you coped with your cancer treatment”) (Helmerhorst, Brake, Warren, Besson, & Ekelund, 2012).

Self-report instruments include diaries, questionnaires, surveys, and interviews. Self-report data may be collected in-person, via telephone, with online and off-line computer-based programs, mobile

phone applications, written documents, or by mail. In global self-reports—that is, measures assessing overall level of intensity or frequency of constructs such as symptoms (e.g., pain), emotions (e.g., fear of pain), beliefs (e.g., catastrophizing), or behaviors (e.g., avoidance behaviors)—individuals are asked to indicate how many times, how often, or how intensive their feelings, thoughts, or actions were.

In self-report measures, it is important to make a decision as to the specific time period for the report (e.g., two weeks). Counting number of cigarettes (e.g., “on average, how many cigarettes do you smoke per day?”) (Shiffman, 2009) is a good example of a behavioral self-report measure. Time line follow-back self-reports gather information on the way an individual has acted, felt, or thought during a pre-selected time period that can cover anywhere from the previous few days up to the previous 12 months. A calendar may be used to structure the report, and personal (e.g., times of leaving for work and returning home; visits at healthcare services) as well as common (e.g., weekends, national holidays) landmarks are used as memory aides to assist recall and engage autobiographical memory (Menon, 1994). End-of-the-day diary refers to number or intensity of symptoms, thoughts, feelings, or acts during the past day.

Self-reports are easy to administer and low cost. However, self-reports are prone to reporting errors and cognitive biases. People may respond to questions with a need for approval or affirmation, or to enhance their self-image. Particularly for older people and those with medical illnesses, there may be cognitive limitations affecting comprehension or recall (Helmerhorst et al., 2012). Self-reports covering longer time periods are more susceptible to memory-related biases including storage and retrieval failures (Menon, 1994). Self-reports in which participants estimate the average occurrence of behaviors, symptoms, thoughts, or feelings may be biased, because instead of engaging in systematic counting, respondents base their answers on broad heuristics such as “digit bias”, that is, the tendency to cluster around rounded values (Shiffman, 2009). For example, people would tend to report that they have missed 10% of prescribed medication rather than reporting missing 12%.

In the past it was often assumed that socially undesirable behaviors, thoughts, or feelings tend to be underreported in self-reports compared to computer-assisted instruments (Brener, Billy, & Grady, 2003). One might expect that computer-based surveys offer a greater sense of privacy, or anonymity compared to a conventional paper-and-pencil survey. However, recent meta-analyses showed that differences in socially desirable outcomes obtained in computer-administered and human-administered self-reports are close to zero ($d = 0.01$; Dodou & de Winter, 2014). This may be a result of more rigorous design applied in recent studies (e.g., random assignment to computer-assisted and paper-and-pencil groups; larger samples) or changes in the use of information technologies

Ecological Momentary Assessment

A critique of global self-report measures is that they tend to miss the effects of daily events and circumstances and are insensitive to changes over time, across contexts, and within the individual. To minimize these issues, the ecological momentary assessment (EMA) approach uses repeated collection of real-time data and micro-processes in acts, thoughts, symptoms, feelings, or experiences (Kuntsche & Labhart, 2013). EMA often uses digital technology, such as smartphones or smart watches to record data, although it originally used paper and pencil diaries (Green, Rafaeli, Bolger, Shrout, & Reis, 2006). Self-reports occur when an individual is prompted to report as the thought or symptom happens, at specific times (respondents are prompted by an alert), or at random times during the day. Repeated assessment (e.g., every two hours) provides an opportunity for evaluating within-subject changes, appearing over time and across contexts (Shiffman, Stone, & Hufford, 2008).

In addition to its strengths, EMA has its weaknesses. If EMA is applied in a population that is undergoing a treatment (either psychological or medical) EMA may interfere with treatment procedures; before choosing this measurement strategy, feasibility and acceptability for the target patient

group should be carefully tested (Shiffman et al., 2008). Compliance with event-oriented protocols of EMA (e.g., feeling pain as a prompt to record) is very difficult to verify. Further, measurement reactivity, observed in research on self-monitoring, is also a limitation of EMA (Shiffman et al., 2008). Individuals who intend to change their acts, thoughts, or feelings may be prone to changes due to the acts of monitoring or constant reminders of the acts, thoughts, or feelings with EMA.

Biomarker-Based Measurement

The last decade has been characterized by an increasing popularity of biomarker- and sensor- based measurement. Biomarkers are objectively measured indicators (e.g., levels of salivary cortisol, heart rate) used to evaluate physiological or biological processes or responses. They are often considered the gold standard, because compared to self-reports they are believed to be less susceptible to reporting biases (memory bias; adherence to treatment protocols). The use of biomarkers allows for obtaining information about clinically significant factors. Biomarker measurement is usually used to provide information about outcomes such as health behavior, adherence, stress, and health (e.g., status or function of immune, cardiovascular, hormonal system; morbidity). There are multiple biomarkers that may be used as the indicators of outcomes or underlying physiological processes. The use of biomarkers also promotes transdisciplinary understanding and highlights the relevance of psychological processes for well-acknowledged health indicators.

Biomarkers of Smoking Behavior

Going back to our example of smoking behavior, an example of a behavioral measurement with a biomarker indicator is the level of expired CO using a breathalyzer, which can provide information about recently smoked cigarettes. The half-life time for CO varies from three to eight hours (Morabia, Bernstein, Curtin, & Berode, 2001); therefore breathalyzers offer an assessment of short-term exposure to tobacco and would be more reliable (that is, less prone to measurement error) than a self-report question. Longer time biomarkers include assessment of cotinine levels in saliva, with half-life of 5–15 days (Morabia et al., 2001).

Biomarkers of Dietary Behavior

Another example of the use of biomarkers is dietary behavior. Vitamin C and carotenoid indicators may be used as biomarkers of dietary behavior and even provide validity for self-report measures. Other biomarkers of dietary behavior are blood plasma markers, which have high validity and are correlated strongly with laboratory-based observations of food intake (Natarajan et al., 2006). Compared to plasma measures, error variance in self-report measures about dietary behavior may be as high as 50% (Natarajan et al., 2006). Unfortunately, biomarker-based measurement of health behaviors often provides little information about behavior patterns changing over time (e.g., across the day). Relying solely on one food biomarker increases the risk of errors. For example, combining plasma levels of vitamin C and carotenoids is essential in research on fruit and vegetable intake: When vitamin C is applied as a biomarker of fruit and vegetable intake, a portion of green pepper is equivalent to 20 portions of carrots, but when carotenoids are used as a biomarker, one portion of carrots is equivalent to more than 45 portions of green peppers (Kuhnle, 2012).

Adherence to Medication

The measurement of adherence to medication conducted with analyses of biological assays of active drug, drug metabolite, or other markers in blood or urine, which confirm active drug ingestion is

usually moderately related to self-report (Lam & Fresco, 2015). Although biomarkers currently may be the most accurate measures of adherence, they have multiple drawbacks. First, an individual's physiological state and metabolism levels affect drug metabolism. Therefore, drug plasma levels differ within and between individuals, even if the same dose of the same medicine was taken (Lam & Fresco, 2015). Second, biomarkers do not allow for revealing patterns of nonadherence across time (Lam & Fresco, 2015). "White coat effect" studies show that patients have higher adherence just before upcoming tests, with a 20% difference in adherence around the time of the visit at healthcare professional compared to one month after the visit (Lam & Fresco, 2015). Moreover, biomarker-based measures involving drawing blood or providing urine samples are considered relatively invasive or intruding respondents' privacy (Morabia et al., 2001), which may diminish participation in research studies and increase attrition.

Stress

Salivary cortisol is one of the best-established biomarkers of acute, chronic, or traumatic stress (Ryan, Booth, Spathis, Mollart, & Clow, 2016). Cortisol is the primary indicator of the function and dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis. However, there are many ways to measure it. There are two state-of-the-art approaches in cortisol studies (Ryan et al., 2016). One involves reactivity to stress that is measured after an experimental exposition to an acute stressor; the second involves the circadian rhythm of diurnal cortisol secretion. The "normal" diurnal slope of cortisol includes a peak at 30–45 minutes after waking up and a steady decline over the hours of the day. Thus, cortisol measurements should be taken upon awakening, 30–45 minutes after awakening, during the day, and in the evening to get a sense of an individual's diurnal pattern. The diurnal profile parameter should account for cortisol awakening response, diurnal slope, and area under the curve index (Ryan et al., 2016). It should be noted that smoking, menstrual cycle, age, and sleep all affect cortisol levels and these must be taken into account.

Salivary measures of cortisol involve swabbing one's own mouth or spitting into a vial. As a result, it is a convenient and low-cost way to measure cortisol. However, there are drawbacks. Collecting salivary cortisol data in this way requires that participants collect cortisol samples at specific times of day, several times per day. In order to have a valid assessment, at least 30 minutes should pass since the participant has consumed food, liquid (other than water), or used tobacco before obtaining each saliva sample. Thus at-home cortisol collection procedures are complex and demanding, even for participants who are healthy and with excellent cognitive functioning.

Inflammation

The most used indicators of the degree of inflammation in the body include cytokines, such as IL-1 or IL-6, tumor necrosis factor- α , C-reactive protein, number of cells denoting an immune response, such as helper-T cells. High levels of inflammation have well-established links with numerous health problems, including cardiovascular diseases, Alzheimer's disease, and frailty (Segerstrom, Out, Granger, & Smith, 2015). However, several inflammatory indicators depend on age and circadian rhythms, and require multiple measurement points (on separate days) to ensure reliability of the measurement (Segerstrom et al., 2015).

Cardiovascular Functioning

Much research in health psychology focuses on the role of cardiovascular risk factors and cardiovascular disease, as well as interventions to reduce mortality and morbidity risk. For example, blood

pressure and heart rate are extensively used in research on the role of stress-related psychosocial factors related to the development or the course of cardiovascular diseases (Segerstrom et al., 2015). However, researchers using these indices struggle between the greater experimental control and precision of these measures obtained in laboratory-based settings and ecologically valid yet “noisy” ambulatory blood pressure (BP) protocols (Segerstrom et al., 2015). Ambulatory BP monitors involve an inflatable cuff worn under clothing and a small control box attached to a waist-belt. Thus, continuous readings over a long period can be obtained. A combination of ambulatory blood pressure protocols with EMA-based self-reports may provide a detailed insight into within-individual processes in which thoughts, feelings, and acts are preceded or followed by changes in cardiovascular function.

Genetic Biomarkers

This is an emerging area within health psychology research (see Chapter 38). Recent theoretical and empirical developments highlight the need for the use of genetic biomarker-based measurement in health psychology research. The accumulation of evidence for associations between better cognitive abilities, social functioning, lower morbidity, and better health resulted in developing theories of system integrity, suggesting a “third factor”—a complex physiological system that fuels these associations (Mottus, Marioni, & Deary, 2015). Quantitative behavioral genetics (i.e., twin and family studies) and candidate gene association studies provide clear evidence for the genetic underpinnings of psychosocial determinants of health and psychosocial health outcomes.

Individual differences in reward-motivated behaviors (eating, substance use) and social function (e.g., bonding) may be explained by effects of dopamine receptor and dopamine transporter genes (Harden, 2014). Individual differences in impulsive behaviors, depression, anxiety, disordered eating, response to environmental stress may be explained by genes regulating serotonin transporter and reuptake (Harden, 2014). In turn, loneliness, partner bonding, perceived marital problems, marital status, anxiety, and depression may be explained by oxytocin and vasopressin receptor genes (Harden, 2014). Recent advances in mapping genetic variations aid the interpretation of genetic association studies, and provide standards on how best to design research and analyze sequencing-based information (Abecasis et al., 2012). Calculating *genome-wide polygenic scores*, which aggregate the effects of thousands of DNA variants from genome-wide association studies, may aid in predicting resilience or risk for distress, behavioral problems, mental health issues, and better or poorer educational achievements (Krapohl et al., 2016).

Sensor-Based Measurement

Technological developments offer sensor-based measurement approach to explaining complex behaviors, such as nutrition, physical activity, or adherence to medication. Sensor-based measurement relies on electronic sensors indicating physical characteristics (e.g., movement, mass) or providing video (e.g., a photo) recordings of an individual and his/her environment.

Adherence to Medication

Medication events monitoring systems (MEMS) include adherence monitoring devices, such as electronic pill containers that register and code information about whether and when the pill was taken (Lam & Fresco, 2015). The basic principle of MEMS is that whenever the medication is removed from the container, a microprocessor embedded in the container records the time and date. It assumes that the patient has taken that specific dose at that particular time (Lam & Fresco, 2015).

MEMS have high validity and reliability as well as high accuracy when compared to biomarker-based indicators (Lam & Fresco, 2015). Furthermore, MEMS can identify whether the nonadherence is sporadic or consistent, to establish medication-taking daily patterns (Lam & Fresco, 2015). These features may make MEMS more useful than biochemical and self-report measures.

Physical Activity

Accelerometers detect acceleration in up to three directions and thus allow determination of the quantity and intensity of movements. Compared to self-reports, triaxial accelerometers have excellent validity and reliability (Van Remoortel et al., 2012). A major limitation of accelerometry involves measuring physical activity among patients moving slowly due to a health condition, such as chronic heart disease, pulmonary diseases, chronic obstructive pulmonary disease, arthritis, or back pain (Van Remoortel et al., 2012). Unfortunately, accelerometers tend to underestimate the total energy expenditure in the case of slow walking, and poorly differentiate between standing and sitting/reclining. Furthermore, accelerometer-based measurement requires continuous wearing of the device on the right hip, for example, for at least 10 hours per day for at least a week, to obtain a reliable and valid assessment of physical activity patterns. Thus, feasibility of measurement is limited for many populations, including children, people with executive function deficits, and those with severe physical disability.

Food Intake and Alcohol Use

Cameras may be used to obtain visual indicators (e.g., pictures) of food consumption and alcohol intake, as well as physical and social environment characteristics. Data may be collected with mobile phone cameras (e.g., research participants are asked to take a picture of each food or drink that they are about to consume) or with wearable cameras (worn around the neck, capturing wide angle images; taking pictures automatically, every 30 seconds or in response to movement, temperature, and light changes; Doherty et al., 2013). Categorizing data obtained with cameras may be time consuming and expensive, as researchers must analyze the content of images visually in order to categorize recorded behaviors or environment (Doherty et al., 2013). The measurement conducted with wearable camera is likely to yield a high amount of missing data, for example, when participants want to protect their privacy or the privacy of other people (Doherty et al., 2013). Emerging research indicates that virtual reality devices may also be useful for health behavior research, but may be prone to many of the same issues.

Step 4. Consider the Psychometric Properties of the Measures

Ensuring that an instrument has good psychometric properties is crucial for forming accurate and replicable research conclusions or a clinical diagnosis. Instruments with good psychometric properties are able to capture the target construct well, do it consistently, and detect even small changes in the target construct that may occur over time (Wasserman & Bracken, 2003). Psychometric characteristics also indicate the level of error in measurement.

Before using any instrument, researchers and practitioners should go through a checklist of psychometric characteristics and consider the pros and cons of the use of an instrument on its own and in comparison to alternatives. This checklist would usually account for the measure's validity, reliability, and sensitivity (Johnston et al., 2016). As there are many excellent books and articles that address these in detail (e.g., Furr, 2018; Wasserman & Bracken, 2003), we will only provide a brief description of each and why it is important.

Validity: What Is Actually Measured

According to the shortest definition, validity is about the meaning of scores (Wasserman & Bracken, 2003). Psychometric validity refers to the extent to which the scores obtained with an instrument exclusively and adequately measure the target construct and guide consequential decision making. An instrument may fail to tap all crucial aspects of the target construct (construct underrepresentation) or account for other (un)related constructs (construct irrelevance). Internal validity of an instrument may refer to tapping the theoretical internal structure of the construct. Internal validity allows for an evaluation if the instrument includes enough diverse content to adequately sample the breadth of relevant domains of the target construct but at the same time the instrument does not lose its coherence and uniformity (Wasserman & Bracken, 2003). External evidence for validity of an instrument is drawn from independent, criterion-related data (e.g., obtained with other instruments; Wasserman & Bracken, 2003). External validity reflects the extent to which the scores indicate the target, independent of actual test performance. Several types of external validity are distinguished (e.g., convergent, discriminant, criterion-related, and consequential).

Reliability: How Well the Measurement Has Been Done

Reliability refers to the level of precision and accuracy of measurement. Reliable instruments allow for measurement that is consistent, accurate, and uniform across testing occasions, across observers, and across samples (Wasserman & Bracken, 2003). Internal consistency would inform about the degree of uniformity and coherence among the parts of the instrument (e.g., its items or sections). Temporal stability refers to the consistency of scores over time. Inter-rater reliability refers to the consistency of variance among independent whereas inter-rated agreement addresses the extent to which the raters make the same ranking.

Sensitivity: What Happens Over Time

Sensitivity refers to the extent to which the instrument may detect small changes occurring over time (cf. Johnston et al., 2016). High sensitivity is vital for detecting changes in longitudinal research or in treatment evaluation. In case of multi-item self-reports health psychologists may consider if the instrument was analyzed applying item-response theory, which takes into account the characteristics of the measure and the respondent to provide information about the sensitivity of each item to detect the underlying construct (Karademas et al., 2016). The evaluation of an instrument may also include signal-detection theory procedures applied to calculate sensitivity of the instrument in terms of its ability to differentiate between information-bearing stimuli and the “noise” of random patterns and confounding variables (Karademas et al., 2016).

Step 5. Consider Implementation of Measurement

Implementation science highlights the need for careful evaluation of the way that health psychology interventions and measurement are applied across and within research or practice settings and contexts (Horodyska et al., 2015). Theories and research show that implementation has immense influence on any measurement process as well as on results of any behavioral medicine intervention, including its effectiveness, safety, and patient burden or satisfaction (Horodyska et al., 2015).

A number of implementation outcomes may be considered when choosing a measurement instrument (Lewis et al., 2015). For example, acceptability of an instrument refers to the perception among implementation stakeholders (e.g., patients/participants, health psychologists,

healthcare system/setting managers) that an instrument is attractive, satisfactory and meets the stakeholders' needs (e.g., in terms of usability, content, complexity, comfort, credibility, delivery). There are at least 50 instruments (including structured interviews, questionnaires) to assess acceptability (Lewis et al., 2015).

The implementation of an instrument may be guided by checklists of critical implementation conditions, some including as many as 83 implementation conditions (Horodyska et al., 2015). These conditions enhance the delivery and effectiveness of any psychosocial actions, in particular those conducted in real-world settings. For example, when using a specific measure, health psychologists should consider how the instrument might affect attrition rates over the course of a study, the simplicity of the instrument, participant burden, or major adaptations for a particular population. One must also consider the training and expertise needed to implement the instrument with fidelity.

Conclusion

Measurement is the backbone of any health psychology study or intervention. The decision to use a particular measurement instrument is a complex process that deserves thought. In this chapter, we suggested ways in which researchers and practitioners can improve the quality of their measurement by using the steps outlined in this chapter.

As measurement is only one of many aspects of research design, health psychologists may tend to make easy and instant decisions guided by two questions: "Does the instrument measure the construct I wanted to measure?" and "Is it reliable and valid?" However, as this chapter shows, choosing the right measure will only strengthen the contribution of a study.

References

- Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., . . . Peterson, J. L. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*, 56–65. doi:10.1038/nature11632.
- Abraído-Lanza, A. F., Cunha Martins, M., & Shelton, R. C. (2015). Breast cancer screening among Dominican Latinas a closer look at fatalism and other social and cultural factors. *Health Education and Behavior*, *42*, 633–641. doi:10.1177/1090198115580975
- Apple, D. K. (2005). *Designing and implementing performance measures handbook*. Lisle, IL: Pacific Crest.
- Apple, D. K. (2006). *Assessment Institute handbook*. Lisle, IL: Pacific Crest.
- Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*, 3186–3191.
- Betancourt, J. R., Green, A. R., Carrillo, J. E., & Ananeh-Firempong, O. (2003). Defining cultural competence: A practical framework for addressing racial/ethnic disparities in health and health care. *Public Health Reports*, *118*, 293–302.
- Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behaviors among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, *33*, 436–457.
- Christie, D. (2015). Assessment in children. In Y. Benyamini, M. Johnson, & E. C. Karademas (Eds.), *Assessment in health psychology* (pp. 278–289). Boston, MA: Hogrefe.
- Conner, M., & Norman, P. (2015). Predicting and changing health behavior: A social cognition approach. In P. Norman & M. Conner (Eds.), *Predicting and changing health behavior* (pp. 1–29). Maidenhead, UK: McGraw-Hill.
- Dodou, D., & de Winter, J. (2014). Social desirability is the same in offline, online, and paper surveys. *Computers in Human Behavior*, *36*, 487–495. doi:10.1016/j.chb.2014.04.005
- Doherty, A. R., Kelly, P., Kerr, J., Marshall, S., Oliver, M., Badland, H., . . . Foster, C. (2013). Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *10*, 22. doi:10.1186/1479-5868-10-22
- Furr, R. M. (2018). *Psychometrics* (3rd ed.). London, UK: Sage.

- Green, A. S., Rafeali, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods, 11*(1), 87–105. doi:0.1037/1082-989X.11.1.87
- Harden, K. P. (2014). Genetic influences on adolescent sexual behavior: Why genes matter for environmentally-oriented researchers. *Psychological Bulletin, 140*, 434–465. doi:10.1037/a0033564
- Helmerhorst, H. J. F., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 103. doi:10.1186/1479-5868-9-103
- Horodyska, K., Luszczynska, A., Hayes, C. B., O'Shea, M. P., Langoien, L. J., Roos, G., . . . Brug, J. (2015). Implementation conditions for diet and physical activity interventions and policies: An umbrella review. *BMC Public Health, 15*, 1250. doi:10.1186/s12889-015-2585-5
- Johnston, M., Benyamini, Y., & Karademas, E. C. (2016). Measurement issues in health psychology. In Y. Benyamini, M. Johnston, & E. C. Karademas (Eds.), *Assessment in health psychology* (pp. 320–334). Boston, MA: Hogrefe.
- Karademas, E. C., Benyamini, Y., & Johnston, M. (2016). Introduction. In Y. Benyamini, M. Johnston, & E. C. Karademas (Eds.), *Assessment in health psychology* (pp. 3–18). Boston, MA: Hogrefe.
- Karp, J. F., Rudy, T., & Weiner, D. K. (2008). Persistent pain biases item response on the Geriatric Depression Scale (GDS): Preliminary evidence for validity of the GDS-PAIN. *Pain Medicine, 9*, 33–43. doi:10.1111/j.1526-4637.2007.00406.x
- Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J. B., Rimfeld, K., von Stumm, S., . . . Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry, 21*, 1188–1193. doi:10.1038/mp.2015.126
- Kuhnle, G. C. (2012). Nutritional biomarkers for objective dietary assessment. *Journal of the Science of Food and Agriculture, 92*, 1145–1149. doi:10.1002/jsfa.5631
- Kuntsche, E., & Labhart, F. (2013). Using personal cell phones for ecological momentary assessment. An overview of current developments. *European Psychologist, 19*, 3–13. doi:10.1027/1016-9040/a000127
- Lam, W. Y., & Fresco, P. (2015). Medication adherence measures: An overview. *BioMed Research International, 2015*, e217047. doi:10.1155/2015/217047
- Leventhal, H., Phillips, L. A., & Burns, E. (2016). The Common-Sense Model of Self-Regulation (CSM): A dynamic framework for understanding illness self-management. *Journal of Behavioral Medicine, 39*, 935–946. doi:10.1007/s10865-016-9782-2
- Lewis, C. C., Fisher, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science and enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science, 12*, e155. doi:10.1186/s13012-015-0342-x
- Luszczynska, A., & Schwarzer, R. (2015). Social cognitive theory. In P. Norman & M. Conner (Eds.), *Predicting and changing health behavior* (pp. 225–251). Maidenhead, UK: McGraw-Hill.
- Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 161–172). New York: Springer.
- Morabia, A., Bernstein, M. S., Curtin, F., & Berode, M. (2001). Validation of self-reported smoking status by simultaneous measurement of carbon monoxide and salivary thiocyanate. *Preventive Medicine, 32*, 82–88. doi:10.1006/pmed.2000.0779
- Mottus, R., Marioni, R., & Deary, I. J. (2015). Markers of psychological differences and social and health inequalities: Possible genetic and phenotypic overlaps. *Journal of Personality, 85*, 104–117. doi:10.1111/jopy.12220
- Natarajan, L., Flatt, S. W., Sun, X., Gamst, A. C., Major, J. M., Rock, C. L., . . . Pierce, J. P. (2006). Validity and systematic error in measuring carotenoid consumption with dietary self-report instruments. *American Journal of Epidemiology, 163*, 770–778. doi:10.1093/aje/kwj082
- Purnell, J. Q., Katz, M. L., Andersen, B. L., Palesh, O., Figueroa-Moseley, C., Jean-Pierre, P., & Bennett, N. (2010). Social and cultural factors are related to perceived colorectal cancer screening benefits and intentions in African Americans. *Journal of Behavioral Medicine, 33*, 24–34. doi:10.1007/s10865-009-9231-6
- Ryan, R., Booth, S., Spathis, A., Mollart, S., & Clow, A. (2016). Use of salivary diurnal cortisol as an outcome measure in randomised controlled trials: A systematic review. *Annals of Behavioral Medicine, 50*, 210–236. doi:10.1007/s12160-015-9753-9
- Segerstrom, S. C., Out, D., Granger, D. A., & Smith, T. W. (2015). Biological and physiological measures in health psychology. In Y. Benyamini, M. Johnston, & E. C. Karademas (Eds.), *Assessment in health psychology* (pp. 227–240). Boston, MA: Hogrefe.
- Shiffman, S. (2009). How many cigarettes did you smoke? Assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment. *Health Psychology, 28*, 519–526. doi:10.1037/a0015197

- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 1*, 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415
- Smith, T. W. (2003). Health psychology. Part Two. Research methods in specific content areas. In I. B. Weiner (Ed.), *Handbook of psychology* (pp. 241–270). Chichester, UK: Wiley.
- Smith, T. W. (2011). Measurement in health psychology research. In H. S. Friedman (Ed.), *The Oxford handbook of psychology* (pp. 42–72). New York: Oxford University Press.
- Van Remoortel, H., Giavedoni, S., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E., . . . Troosters, T. (2012). Validity of activity monitors in health and chronic disease: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 9*, e84. doi:10.1186/1479-5868-9-84
- Vetter, T. R., Bridgewater, C. L., & McGwin, G. (2012). An observational study of patient versus parental perceptions of health-related quality of life in children and adolescents with a chronic pain condition: Who should the clinician believe? *Health and Quality of Life Outcomes, 10*, e85. doi:10.1186/1477-7525-10-85
- Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology. Volume 10: Assessment psychology* (pp. 43–66). Hoboken, NJ: John Wiley & Sons Inc.