

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



LSHTM Research Online

Borghì, J; Lohmann, J; Dale, E; Meheus, F; Goudge, J; Oboirien, K; Kuwawenaruwa, A; (2017) How to do (or not to do)... Measuring health worker motivation in surveys in low- and middle-income countries. Health policy and planning. ISSN 0268-1080 DOI: <https://doi.org/10.1093/heapol/czx153>

Downloaded from: <http://researchonline.lshtm.ac.uk/id/eprint/4645597/>

DOI: <https://doi.org/10.1093/heapol/czx153>

**Usage Guidelines:**

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

<https://researchonline.lshtm.ac.uk>

## Box 1 : Overview of Motivation Theories

Theory	Original source	Brief description of the theory
<b>Content theories</b>		
	The Need Hierarchy	Maslow, 1943  Maslow's hierarchy of needs is often portrayed in the shape of a pyramid where physiological needs are at the bottom of the pyramid and considered to be most fundamental. These are followed by safety, then love and belonging, which are followed by esteem and finally the need for self-actualization at the top. When applied to work motivation, it implies that physiological needs (such as salary, secure place to work) should be satisfied before anything else.
	ERG Theory: Existence, Relatedness, Growth	Alderfer, 1972  Developed out of the Maslow's hierarchy of needs. Existence relates to a person's physical needs such as food, clothing, and shelter, similar to Maslow's physiological and safety needs. Relatedness is concerned with the desire people have for maintaining important interpersonal relationships. Growth relates to a person's needs of personal development. Unlike Maslow's theory, lower level need does not necessarily have to be gratified for a higher level to become relevant. This implies that in a workplace managers must recognize their employees' multiple simultaneous needs.
	Two-Factor Theory: Motivators vs. Hygiene Factors	Hertzberg, 1959  Basic idea is that factors which lead to satisfaction such as achievement, intrinsic interest in the work, and involvement in decision making, are distinct from those which lead to job dissatisfaction, such as working conditions, salary, and administrative practices.
	Learned Need Theory: Need for Achievement, Need for Power, and Need for Affiliation	McClelland, 1976  According to McClelland, all humans have three motivators: a need for achievement, a need for affiliation, and a need for power. However, there is one dominant motivator, which is acquired ("learned") through life experience and culture. People with different dominant motivators have different characteristics appropriate for different types of job and positions.
<b>Process theories</b>		
	Equity Theory	Adams, 1965  Focuses on outcomes, a person's perception of fairness as a motivator. It introduced the concept of social comparison where motivation is based on what a person considers to be fair when compared to others. Employees who perceive inequity when comparing themselves to others in the organization will seek to eliminate it by altering inputs or outputs.

	Expectancy Theory: Job Outcomes, Valences, Instrumentality, and Expectancy	Vroom, 1964	Defined as an action-outcome estimate: people choose their behaviors (effort level) based on their perceptions of whether the behavior is likely to lead to valued outcomes.
	Reinforcement Theory or Operant Conditioning: Stimulus, Response and Consequence	Skinner, 1969	Behavior as a “function of its consequences”, desirable behavior can be increased through rewards or reinforcement techniques. Reinforcers can be financial or non-financial (i.e. informational).
	Cognitive Evaluation Theory (CET): Intrinsic and Extrinsic Motivation	Porter and Lawler, 1968; deCharms, 1968	Building on Vroom’s (1964) theory of motivation, Porter and Lawler (1968) proposed a model of intrinsic and extrinsic work motivation, where it appeared that contingent, tangible rewards and other extrinsic factors such as competition and evaluations could undermine intrinsic motivation. Basic assumption in CET is that people have an innate need to feel autonomous and competent, and contingent rewards could undermine these feelings.
	Goal Setting Theory	Locke and Latham, 1984	People’s actions are driven by goals, they exert more effort when they have specific goals which are difficult but are seen as attainable. Goals need to be accepted, hence the importance of the goal setting process.
	Social Cognitive Theory (self-efficacy)	Bandura, 1977	Belief in one’s capabilities to successfully execute the behavior, which is needed for a particular task. Experiments showed that even holding abilities constant, people who were more confident exerted more effort, persisted longer and performed better at a task than those who had less confidence.
	Self-Determination Theory	Deci and Ryan, 2002	Expands on CET, moving away from a simple dichotomy of intrinsic vs. extrinsic motivation. It characterizes extrinsic motivation as a continuum where there are many “types” of extrinsic motivation which differ in their degree of autonomy and internalization. Between amotivation and intrinsic motivation, along a continuum, there are four types of extrinsic motivation, with external being the most controlled type of extrinsic motivation, and introjected, identified, and integrated being progressively more self-determined or autonomous.

Sources: Shortell and Kaluzny, 2006 ; Mitchell, 1997.

## **BOX 2: Defining Concepts**

**Construct:** used to refer to motivation as a theoretical concept.

**Dimension:** Dimensions refer to sub categories of motivation when motivation is conceptualised as being multi-dimensional. For example, extrinsic and intrinsic motivation may be identified as distinct sub-dimensions of motivation. Dimensions can be predefined based on theory and literature, or emerge during the research process, e.g. if there is no clear understanding of the construct in a specific context yet, or a combination of both.

**Factor:** The term 'factor' used in factor analysis language to refer to the unobservable (or latent) dimension in question, which is measured by the items pertaining to it. For instance, the unobservable factor 'intrinsic motivation' might be indicated/measured by five directly observable items (in the sense that respondents give observable answers to these items).

**Item:** refers to a statement or question in a survey tool to measure motivation.

**Response scale:** refers to the response options presented to the respondent in relation to an item (survey statement/question). The term is used especially when there are multiple, ordered response options such as with a Likert scale.

**Measurement scale (or just 'scale'):** refers to a set of items intended to measure the same construct (e.g. motivation)

**Survey:** refers to the entire questionnaire which usually includes more than one measurement scale and a variety of other questions (demographic, work related).

### **BOX 3: Finding the right terms**

Focus group discussions and interviews are an important way of identifying the appropriate way of communicating concepts related to motivation and eliciting meaningful responses in local language. Constructs such as motivation are not easily and directly translated and understood in the same way across cultures. In Afghanistan, two focus group discussions with seven health workers in each from two different facilities in Kabul were conducted exploring general questions on reasons for choosing the profession and attitudes towards work. We realised through this process that the word most frequently used by respondents in these discussions when referring to motivation in Dari was the word for encouragement, *tashweeq*. In Dari, *tashweeq* means "evocation of *shauq*", *shauq* being a word for desire, zeal, or inclination. When it is combined with the verb "*kardan*" (to do), *tashweeq kardan*, it means to encourage. However, when combined with the verb "*shodan*" (to become), *tashweeq shodan*, it might acquire an intrinsic aspect. However, in Persian, the word "*angize*" means motivation. However, "*angize*" was used more rarely, although it seemed to be understood by all health workers (Dale, 2014). This is very similar to findings of the qualitative study in Kenya and Benin (Mathauer, Imhoff, 2006), according to which over 50% of health workers in Benin equated motivation with prospective "encouragement" with one fourth of these explicitly mentioning financial encouragement, while another 40% considered "being motivated" as having the necessary means to work and get recognition. As the authors of this study put it, the majority understood motivation as "an incentive, and not as a state of mind" (Mathauer, Imhoff, 2006). In Tanzania, respondents differentiated between: motivation as a desire to serve "*kuwa na moyo*"; and motivation driven by monetary benefits, social recognition, power "*motisha*".

This type of qualitative analysis through focus group discussions with the target group will play an important role in refining one's conceptualization of motivation, developing items for the questionnaire, and finding a way of eliciting the appropriate construct. This will be context specific and may need to be adapted to different respondent types, particularly if there are large differences in the level of education (e.g. doctor versus community health worker).

#### Box 4: Question Format

There are different ways of structuring questions in motivation surveys. When assessing motivational composition, researchers typically use a series of items, against which respondents have to rate their level of agreement. These can be phrased as affirmations or in the negative. For example when examining the relevance of staffing to motivation researchers could use the following item: 'there are enough providers at this facility'. This could be phrased using negative wording: 'there are not enough providers at this facility'; or could be presented in the affirmative, but conveying a negative concept: 'there are insufficient providers at this facility'. The use of negative wording, or negative concepts in the affirmative may help to reduce acquiescence bias (Prytherch et al. 2012) but can sometimes confuse respondents, and their responses to equivalent positive items do not always correspond. In Burkina Faso, for instance, JL presented respondents with the following two items (spaced, with other items in between):

- 1) "I wish I worked in another health facility" and
- 2) "I wouldn't want to work in another health facility".

Although equivalent in meaning, the following means were obtained on a 0-10 response scale: 1) 5.2; 2) 3.0 (non-reversed). Respondents' answers to the two questions correlated at only -.24.

Items should be kept simple, avoiding leading or double-barreled questions. They can be assessed for readability using the Flesch reading ease formula using Microsoft Word. It is recommended that items should not exceed the reading level of 6<sup>th</sup>-7<sup>th</sup> grade. It is also recommended to check understanding among a representative sample of health workers, as education levels vary by setting (Dale 2014).

Along with the development of items, researchers must also choose a response scale. The responses to items can take many forms (Streiner et al. 2008). Options include dichotomous response options (e.g. yes-no, true-false, agree-disagree) or rating scales. The Likert rating scale is widely used in the literature on health worker motivation (e.g. (Chandler et al. 2009; Weldegebriel et al. 2016) (Alhassan et al. 2013)), and preferred by the authors of this paper. Considerations here include the number of response categories, and whether or not to include a neutral category. Too few categories will result in a loss of information and less variance in data, impacting reliability. A larger number of categories will make the tool more sensitive to detecting a shift in motivation levels if that is a research aim. However, too many categories might overwhelm respondents and put into question meaningfulness of differences between categories. The number of categories should be commensurate with respondent ability to discriminate which will vary by context and numeracy (Preston & Colman 2000). Five to nine categories are considered in most circumstances (Streiner et al. 2008). In Afghanistan, the team used 5 categories (Dale 2014). In Tanzania, the team opted for 4 response categories as respondents had difficulty processing more than that. In Burkina Faso, the team used 11 categories with a visual aid (depicting cubes as a visual representation of the 'amount' of agreement) as a compromise between the team's need for sufficient variance in data, respondents' processing capacity, and local 'positivity norms' leading respondents to primarily consider the positive end of the response scale regardless of their actual sentiments. However, the 11 categories might have been overwhelming to respondents, so that the team would recommend 7-9 categories in future research (Lohmann et al. 2017).

A number of response options can be associated with the Likert scale for motivation measurement. 'Agreement-disagreement' response options are often used when measuring motivation through proxies. Response options relating to the 'importance' or the 'frequency' of a given item may also be

appropriate in certain contexts. It is important to ensure the response options are consistent with the associated items. For example, when enquiring about the work environment, you could present the question as follows, using a 4 point Likert scale: *This facility is well stocked with drugs and supplies. Strongly agree, agree, disagree, strongly disagree.*

Items and response options should be chosen so as to be meaningful in a given context. In Afghanistan, the team presented items such as “I work in this job because I have a chance to help other people through my work”, asking respondents to indicate their degree of agreement (Dale 2014). The research in Burkina Faso found respondents to have difficulties with the abstraction processes required to answer such items, and instead presented a list of reasons why people might be motivated to work in their job, asking respondents to indicate to what extent these were important to them personally (Lohmann et al. 2017). It is not a problem to include different response options within the same questionnaire. However, it is advisable to use the same response format (both in terms of number of categories and response options) for a set of items pertaining to the same construct (i.e. if intrinsic motivation is measured with 5 items) so that comparability of responses to items measuring the same construct is preserved. The advantage of response options that relate to ‘satisfaction’ rather than ‘agreement’ is that they can be communicated in conversational language, and do not present the respondent with a pre-formulated response (which may be leading). Though this formulation would not be appropriate for direct measures of motivation.

### **BOX 5: Survey Administration**

Self-administered surveys are often seen as preferable as they maximize perceived confidentiality, thus minimizing social desirability biases, and allow respondents to choose the time of response. However, they are prone to misunderstanding of instructions and acquiescence bias ('rushing through' e.g. by always answering 'yes' or giving the highest/lowest score), and tend to have relatively low response rates without substantial follow-up efforts by the researcher. They also require participants to be able to read and write, which is usually not a concern with skilled healthcare personnel, but can be with auxiliary personnel or community health workers.

For these reasons, interviewer-administered face-to-face surveys are much more common in public health research in LMICs. The risk for social desirability bias in face-to-face surveys can be minimized by training interviewers to be sensitive to the private nature of interview questions, by repeatedly reminding respondents of the confidentiality of their answers, and by administering the survey in a setting that maximises privacy.

In Burkina Faso, JL opted for a 'hybrid approach' to combine the advantages of both forms of administration, minimizing the risk for response biases while at the same time maximizing response rates and data quality. Interviewers read out instructions and all items as in a face-to-face interview. However, respondents were given a separate questionnaire copy to read along and enter their own answers privately rather than disclosing to the interviewer (Lohmann et al. 2017).



### **BOX 6: What is the difference between PCA and factor analysis?**

Although often used interchangeably, PCA and factor analysis are conceptually rather different. PCA is primarily a data reduction [technique used to create indices \(or reduce a number of variables into a single index\)](#). Factor analysis is an umbrella term referring to different techniques aiming to relate the underlying unobservable construct(s) ('latent variable', 'factor') to be measured, i.e. a respondent's level of motivation, to the items intended to measure it. Two general types of factor analysis are distinguished: exploratory factor analysis (EFA), where there is no a priori theory about which items measure which factor and the researcher derives the 'factor structure' (number of factors, e.g. different motivation dimensions; item-factor assignment) of the scale from the data; and confirmatory factor analysis (CFA), where the researcher has a priori hypotheses of the scale's factor structure and examines whether these hypotheses consistent with the data. Within EFA there are two main statistical methods for factor extraction: principle axis factoring (PAF) (the default method used by Stata when using the 'factor' command) and principal component analysis (PCA, with rotation) (Costello & Osborne 2005). Structural equation modelling is the standard statistical technique for CFA and can also be used for EFA. However, alternative data extraction methods exist for both types of factor analysis and might be more appropriate depending on the data (Costello & Osborne 2005).

