RESEARCH · RESEARCH DESIGN

From Research Question to Hypothesis: Operationalising Concepts

How abstract concepts such as loyalty, trust or quality are turned into measurable variables and testable hypotheses, the main decisions involved, and the mistakes that most often weaken the link between concept and measure.

ModQu research article · 2,233 words · 21 references (APA 7th edition)

Research questions are usually written in the language of concepts: customer loyalty, brand trust, employee engagement, service quality. Data, however, consist of specific observations: a rating on a scale, a purchase record, a count of complaints. The process that connects the two is called **operationalisation**. It turns abstract concepts into variables that can be measured, and turns broad research questions into hypotheses that can be tested. Operationalisation is where many studies quietly succeed or fail. A study can be perfectly designed and analysed and still mislead if its measures do not capture the concepts its conclusions are about. This article explains how operationalisation works, the main decisions it involves, and how to judge whether it has been done well.

From concept to measure

The idea that concepts should be defined by the operations used to measure them is associated with the physicist Bridgman (1927), who argued that a concept such as length means nothing more than the set of operations by which length is measured. Strict operationism of this kind was soon criticised, because it implies that two different measuring procedures define two different concepts, so that intelligence measured by one test would be a different thing from intelligence measured by another. Cronbach and Meehl (1955) offered a more flexible view. A concept, or construct, is defined by its place in a network of theoretical relationships, and any particular measure is an imperfect indicator of it. Several measures can indicate the same construct, and the construct is more than any one of them.

The sociologist Paul Lazarsfeld (1958) described the path from concept to measure in four steps, which remain a useful guide.

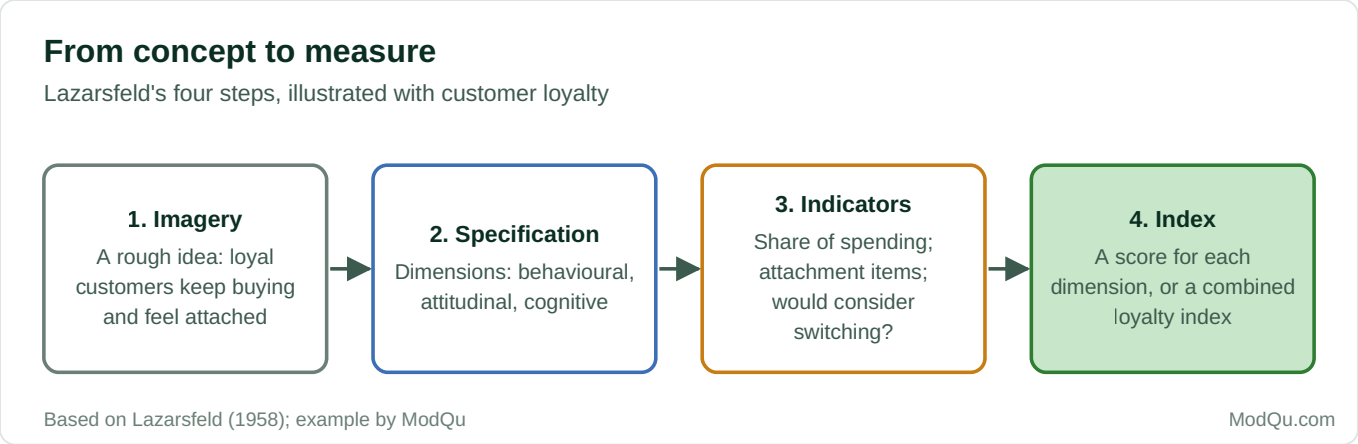


Figure 1. Lazarsfeld's four steps from concept to measure, illustrated with customer loyalty. Based on Lazarsfeld (1958); example by ModQu.

Imagery. The researcher begins with a rough idea of the concept, often drawn from theory or observation. Customer loyalty, for instance, suggests repeated purchase and some kind of attachment.

Specification. The concept is broken into its dimensions. Loyalty might have a behavioural dimension (what customers do), an attitudinal dimension (how they feel) and a cognitive dimension (whether they would consider alternatives).

Selection of indicators. For each dimension, the researcher chooses observable indicators. Behavioural loyalty might be indicated by the share of category spending with the brand; attitudinal loyalty by agreement with statements such as "I feel attached to this brand"; cognitive loyalty by whether customers would consider switching.

Formation of indices. Finally, the indicators are combined into measures that can be analysed, such as a score for each dimension or an overall index.

Adcock and Collier (2001) proposed a similar framework that distinguishes the background concept (the broad range of meanings a term carries), the systematised concept (the specific definition adopted in a study), the indicators, and the scores. Each step involves choices, and measurement validity depends on whether each step faithfully carries the meaning of the one before.

Defining the concept

The first and most neglected step is a clear conceptual definition. MacKenzie et al. (2011) argued that poorly defined constructs are among the most common weaknesses in management and information systems research. A good definition specifies what the construct is, what it is not, the entity it applies to (a person, a team, an organisation) and whether it has distinct dimensions. Without such a definition, there is no standard against which to judge whether a measure is valid.

Rossiter (2002) went further, arguing that the definition of a marketing construct should specify the object (such as a brand), the attribute (such as quality) and the rater (such as customers), and that measures should be judged primarily by whether their content matches that definition, rather than by statistical properties alone.

Choosing indicators

Once the concept is defined, the researcher must choose indicators. Several considerations apply.

Content coverage. The indicators together should cover the whole of the concept as defined. A measure of service quality that asks only about speed omits courtesy, accuracy and reliability.

Level of measurement. Stevens (1946) distinguished four levels at which variables can be measured: nominal (categories with no order), ordinal (ordered categories), interval (equal intervals but no true zero) and ratio (equal intervals and a true zero). The level determines which statistical operations are meaningful. An average of a nominal variable, such as region coded 1 to 5, is meaningless; an average of ratings on an ordinal scale is common practice but rests on assumptions that should be acknowledged.

Multiple indicators. A single indicator is vulnerable to the quirks of its wording and format. Webb et al. (1966) argued for multiple operationism: measuring a concept in several different ways, so that the weaknesses of each are offset by the others. If different measures agree, confidence in the concept grows; if they disagree, the disagreement is itself informative. Shadish et al. (2002) listed mono-operation bias, the reliance on a single operationalisation, as a threat to construct validity.

Existing measures. Where validated measures exist, using them saves effort and allows comparison with other studies. But validation in one setting does not guarantee validity in another, and measures should be checked in the population being studied.

Reflective and formative indicators

An important but often overlooked decision concerns the direction of the relationship between a construct and its indicators.

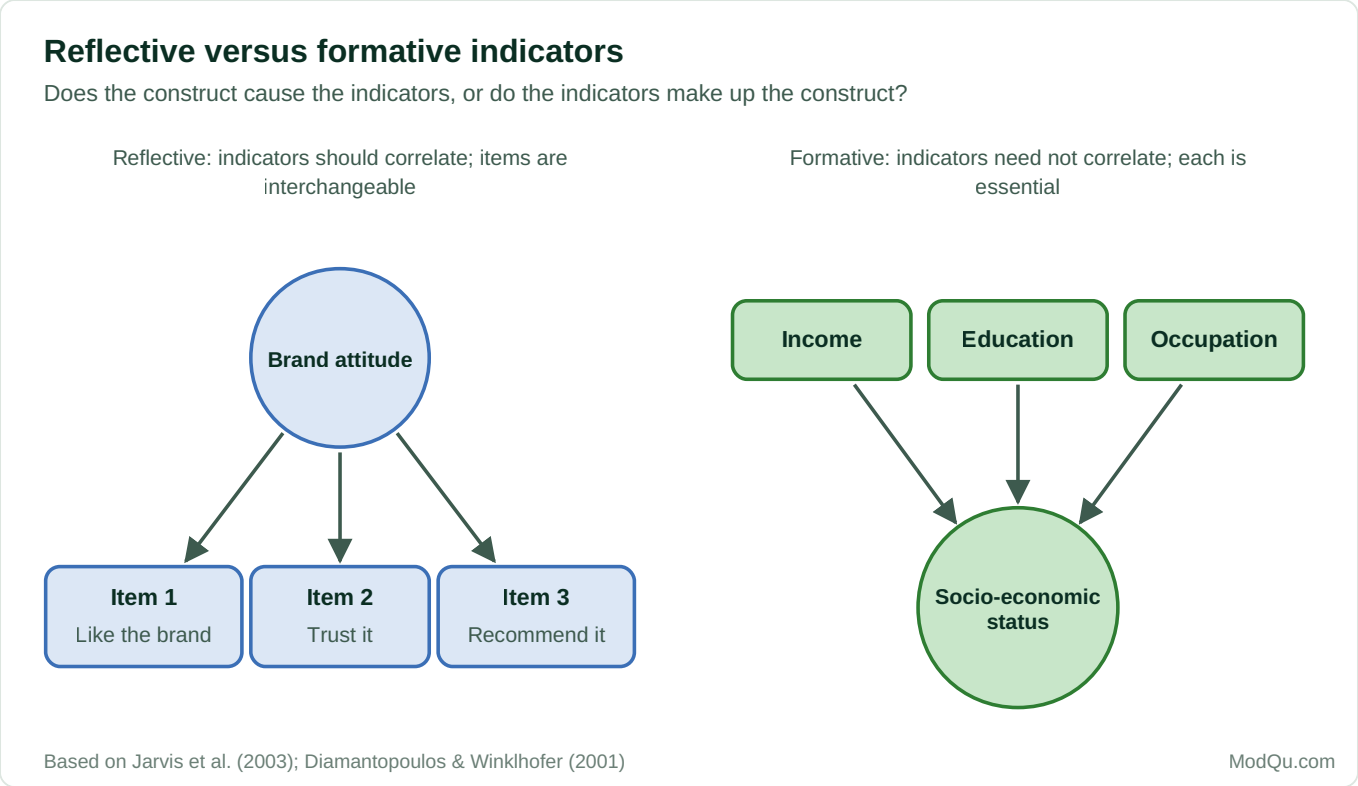


Figure 2. Reflective and formative measurement. The direction of the arrows determines which validation methods are appropriate. Based on Jarvis et al. (2003) and Diamantopoulos and Winklhofer (2001); diagram by ModQu.

In a **reflective** model, the construct causes the indicators. A person's underlying attitude towards a brand produces their answers to several questions about it, so the answers should be correlated, and any one indicator could be dropped without changing the meaning of the construct. In a **formative** model, the indicators cause or make up the construct. Socio-economic status is formed by income, education and occupation; these need not be correlated, and dropping one changes what the index means.

Diamantopoulos and Winklhofer (2001) set out procedures for building formative indices, and Jarvis et al. (2003) reviewed published marketing research and found that a substantial share of constructs had been modelled as reflective when they should have been formative. The error matters because the statistical tests used to validate reflective scales, such as checks of internal consistency, are inappropriate for formative indices and can lead researchers to discard essential components.

Developing new measures

When no suitable measure exists, a new one must be developed. Churchill (1979) proposed a widely used procedure for marketing constructs: specify the domain of the construct, generate a pool of candidate items, collect data, purify the measure by removing poorly performing items, collect new data, and assess reliability and validity. Hinkin (1998) provided a similar step-by-step guide for survey measures in organisational research. MacKenzie et al. (2011) integrated these and later developments into a ten-step process, from conceptual definition through to the development of norms.

These procedures share a principle: measures should be developed iteratively and tested in data, not simply written and assumed to work.

From concepts to hypotheses

A hypothesis states an expected relationship between operationalised concepts. Turning a research question into a hypothesis requires specifying the variables, the direction of the expected relationship and, ideally, its form or size.

Consider the question: does a loyalty programme increase customer loyalty? Operationalised, it might become: customers who join the loyalty programme will increase their share of category spending with the retailer over the following six months, relative to comparable customers who do not join. This version names the independent variable (programme membership), the dependent variable (share of spending), the comparison, the direction and the time frame.

Popper (1959) argued that the more a hypothesis forbids, the more it says. A hypothesis that predicts only "some relationship" is weak. Meehl (1967) warned that in the social sciences, where almost every variable is correlated with almost every other to some degree, predicting a non-zero relationship is a very weak test, since with a large enough sample it will nearly always be confirmed. Stronger hypotheses predict a direction, a size or a pattern that a rival explanation would not predict.

Operationalisation in qualitative research

Operationalisation is usually discussed in quantitative terms, but qualitative research faces the same underlying problem: connecting the concepts in a research question to what is actually observed. A qualitative study of trust between suppliers and buyers must still decide what will count as evidence of trust. Will it be participants' own use of the word, descriptions of behaviour such as sharing confidential information, or the absence of formal contracts? Adcock and Collier (2001) argued that measurement validity is a shared standard for qualitative and quantitative research, because both must show that their observations correspond to the concepts they claim to describe.

The difference is that qualitative operationalisation is often iterative and transparent rather than fixed in advance. Researchers may begin with sensitising concepts that guide what they look for, then refine their definitions as they learn how participants understand the phenomenon. Good practice is to document these refinements, so that readers can see how the concepts used in the findings relate to the evidence. Coding frames, with definitions and examples for each code, serve the same purpose as a measurement protocol in a survey.

Checking the measures once data are in

Operationalisation does not end when data collection begins. Once data are available, researchers can check whether the measures behave as the conceptual definitions imply. Indicators of a reflective construct should correlate with each other more strongly than with indicators of different constructs. Measures should relate to other variables in the ways that theory predicts, which Cronbach and Meehl (1955) described as testing the construct's nomological network. Groups that

should differ on the construct, such as long-standing customers and those who have just cancelled, should in fact differ. When these checks fail, the problem may lie with the measure, the definition or the theory, and the honest response is to report the failure rather than adjust the measure until it passes.

Common problems

Measures that drift from the concept

Flake and Fried (2020) described "questionable measurement practices": undocumented decisions about how constructs are measured, such as altering scales, dropping items or combining measures in unreported ways. These practices make findings hard to evaluate and can inflate false positive results, since a researcher who tries several versions of a measure can report whichever works best. Simmons et al. (2011) showed how such flexibility, combined with other analytic choices, can make almost any hypothesis appear supported.

A single number standing in for a concept

Business research is especially prone to adopting a single metric as if it captured a whole concept. The Net Promoter Score, based on one question about likelihood to recommend, was promoted as the single number a company needs to grow (Reichheld, 2003). Keiningham et al. (2007) tested this claim with longitudinal data and found that the score did not outperform other measures of satisfaction and loyalty in predicting growth. A single indicator may be convenient, but it rarely captures every dimension of a concept such as loyalty.

Measures that become targets

When a measure is used to reward or manage people, it can lose its link to the concept it was meant to indicate. Campbell (1979) observed that the more any quantitative social indicator is used for social decision-making, the more it will be subject to corruption pressures and the more apt it will be to distort the processes it was intended to monitor. A call centre judged on call length may shorten calls without improving service. Operationalisation should anticipate how a measure might be gamed if it becomes a target.

Confusing the measure with the concept

Borsboom et al. (2004) argued that a test is valid only if the attribute it measures exists and causes variation in the scores. If a construct is really a loose collection of different things, no single scale can measure it well. Researchers should be prepared to discover that a concept needs to be split into several, or that a measure captures something different from what was intended.

A practical checklist

Before collecting data, it is worth confirming that:

1. each key concept has a written definition stating what it includes and excludes;

2. the dimensions of each concept have been identified;
3. each dimension has at least one, and preferably several, indicators;
4. the level of measurement of each variable is known and suits the planned analysis;
5. the direction of the construct–indicator relationship (reflective or formative) has been considered;
6. existing measures have been checked for validity in the population studied, or new measures piloted;
7. hypotheses state the variables, direction, comparison and time frame; and
8. measurement decisions are documented, so that changes made later can be reported.

Conclusion

Operationalisation is the bridge between the concepts in a research question and the numbers or observations in a dataset. Each step across the bridge, from imagery to specification, indicators and indices, involves choices that determine whether the findings are about what the researcher claims. Clear definitions, multiple and appropriate indicators, attention to how indicators relate to their constructs, and hypotheses that make risky predictions are the safeguards. When they are in place, a study's conclusions can be read in the language of its concepts. When they are missing, the conclusions may be precise, but they may be about something else.

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review, 95*(3), 529–546. https://doi.org/10.1017/S0003055401003100

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bridgman, P. W. (1927). *The logic of modern physics*. Macmillan.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64–73. https://doi.org/10.1177/002224377901600110

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218. https://doi.org/10.1086/376806

Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, *71*(3), 39–51. https://doi.org/10.1509/jmkg.71.3.039

Lazarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus*, *87*(4), 99–130.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293–334. https://doi.org/10.2307/23044045

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Popper, K. R. (1959). *The logic of scientific discovery*. Hutchinson.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*(12), 46–54.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*(4), 305–335. https://doi.org/10.1016/S0167-8116(02)00097-6

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally.