RESEARCH · PHILOSOPHY AND PARADIGMS

# Theory-Building and Theory-Testing Research

How theory-building and theory-testing research differ in purpose, logic and design, how to judge each, and why the strongest research programmes move between the two.

ModQu research article · 1,556 words · 16 references (APA 7th edition)

Theory sits at the centre of research, but studies relate to it in different ways. Some studies set out to **build** theory: to identify new concepts, propose relationships between them, and explain a phenomenon that existing ideas do not cover well. Others set out to **test** theory: to derive predictions from an established explanation and check them against data. The two activities call for different questions, designs and standards of evidence. Confusing them leads to studies that do neither well, such as an "exploratory" project that claims to have confirmed a hypothesis, or a "test" whose hypothesis was invented after the data were analysed. This article explains the difference, describes how each kind of study is designed and judged, and shows how they fit together.

## What counts as theory?

Before distinguishing building from testing, it helps to be clear about what a theory is. Whetten (1989) proposed that a theory answers four questions: what factors are involved, how they are related, why the relationships hold, and who, where and when the explanation applies. The "why" is essential. A list of variables, a diagram of boxes and arrows, or a set of hypotheses is not a theory unless it also explains the mechanism that links them. Sutton and Staw (1995) made the same point in reverse, listing what theory is not: references to other people's work, data, lists of variables, diagrams and hypotheses can all support a theory, but none of them is a theory on its own.

Bacharach (1989) offered criteria for evaluating theories. A good theory must be falsifiable, meaning that it could be shown to be wrong, and it must have utility, meaning that it explains and predicts something worth knowing. These criteria apply whether a study is building a theory or testing one, but they are applied at different stages.

## Theory-building research

Theory-building studies begin where existing explanations are weak, absent or contradictory. Their aim is to develop new concepts and propose how they are related, so that later studies can test the resulting propositions.

The logic of theory building is mainly inductive and abductive. Researchers gather rich data, often qualitative, and look for patterns that suggest new concepts or relationships. Glaser and Strauss (1967) set out grounded theory as a systematic approach to building theory from data, in contrast to the then dominant practice of testing theories derived from "grand" sociological frameworks. Eisenhardt (1989) described a procedure for building theory from multiple case studies: selecting cases that are likely to extend emerging ideas, collecting data from several sources, comparing cases to find

patterns, and shaping propositions that are then compared with existing literature. Gioia et al. (2013) developed a structured approach that moves from participants' own terms, through researcher-defined themes, to broader theoretical dimensions, presenting the result as a data structure that makes the path from evidence to concept visible.

Theory building is also an act of imagination. Weick (1989) described it as "disciplined imagination": researchers generate many possible explanations and select among them using criteria such as plausibility, interest and fit with the data. Timmermans and Tavory (2012) argued that the key moment is the encounter with a surprising finding that existing theory cannot explain, which prompts an abductive leap to a new explanation.

The products of theory building are concepts, frameworks and propositions. They are judged by their originality, their usefulness and the credibility of the evidence and reasoning behind them. Corley and Gioia (2011) argued that a theoretical contribution should be original, in the sense of offering an insight that is not obvious from what is already known, and useful, in the sense of advancing scientific understanding or informing practice.

## Theory-testing research

Theory-testing studies begin with an established explanation and ask whether its predictions hold. The logic is mainly deductive: hypotheses are derived from the theory, and data are collected to see whether the predicted relationships appear.

Testing calls for designs that can rule out alternative explanations. Experiments, with random assignment to conditions, are the strongest design for testing causal claims, and quasi-experimental and longitudinal designs offer weaker but often more feasible alternatives (Shadish et al., 2002). Measurement must be valid and reliable, and the analysis should be specified before the data are seen, so that the test is a genuine risk to the theory.

Two familiar theories illustrate how testing works. The theory of planned behaviour proposes that intentions, shaped by attitudes, social norms and perceived control, predict behaviour (Ajzen, 1991). The technology acceptance model proposes that perceived usefulness and ease of use predict whether people adopt a new technology (Davis, 1989). Each produced clear, measurable predictions, and each has been tested in a very large number of studies across settings. That accumulated testing is what has established the range of conditions in which each theory works well and the conditions in which it does not.

A theory-testing study is judged by the severity of its test. Meehl (1978) warned that many tests in the social sciences are weak because they predict only that some relationship will be non-zero, which is almost always true with a large enough sample. A severe test predicts a specific direction and size of effect, or a pattern that a rival theory would not predict.

## Fitting the design to the purpose

Edmondson and McManus (2007) argued that good research depends on **methodological fit**: consistency between the state of prior theory, the research question, the data collected and the analysis. When theory in an area is nascent, open-ended qualitative research that builds new constructs fits best. When theory is mature, with well-established constructs and measures, quantitative testing of precise hypotheses fits best. In between, where theory is intermediate, mixed

designs that combine new constructs with established ones are often most productive. Studies that ignore this, for example by testing hypotheses about constructs that have never been properly defined, or by running exploratory interviews in a field with mature measures, tend to produce weak contributions.

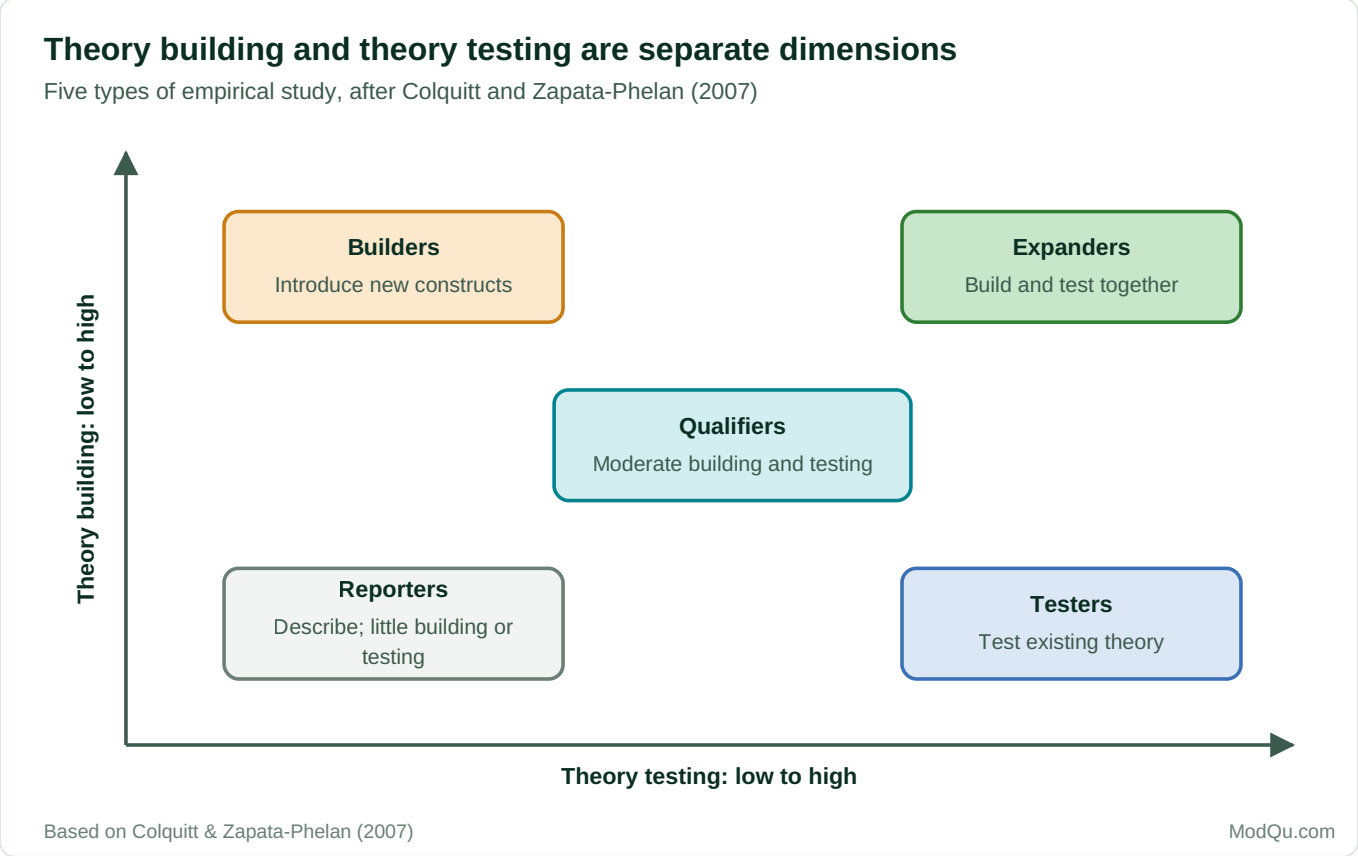



Figure 1. Building and testing as two separate dimensions, with the five types of empirical study identified by Colquitt and Zapata-Phelan (2007). Diagram by ModQu.

## How much of each?

Colquitt and Zapata-Phelan (2007) rated every empirical article published in the *Academy of Management Journal* over five decades on two separate scales: how much it built new theory, and how much it tested existing theory. They grouped articles into five types: reporters, which did little of either; testers, which mainly tested existing theory; qualifiers, which did moderate amounts of both; builders, which mainly introduced new ideas; and expanders, which did a great deal of both. Over time, the share of articles that both built and tested theory rose, and articles that did so tended to be cited more. Their study suggests that building and testing are not alternatives on a single scale but two separate dimensions, and that the most influential work often scores highly on both.

## Building and testing in sequence

Theories rarely move from building to testing in a single step. A typical sequence starts with qualitative work that identifies a phenomenon and proposes concepts to describe it. Measurement work then develops instruments that capture those concepts reliably. Early quantitative studies test whether the proposed relationships appear at all, often in convenient samples. Later studies test the theory more severely, in representative samples, over time, against rival explanations and in experiments where possible. At each stage, anomalies may send researchers back to building, refining the concepts or adding new ones.

The technology acceptance model illustrates this cycle. Davis (1989) began by developing and validating measures of perceived usefulness and ease of use, drawing on earlier theory, before testing whether they predicted intentions to use a system. Later researchers extended the model with additional constructs when tests revealed its limits in particular settings. The theory therefore grew through alternating periods of building and testing, rather than through either one alone.

## Common problems

Several problems recur when the distinction is blurred.

**Presenting exploration as confirmation.** When hypotheses are formed after looking at the data but reported as if they had been predicted, the result looks like a test but is not one. Kerr (1998) called this HARKing, hypothesising after the results are known. The remedy is to label exploratory findings honestly and test them in new data.

**Testing before the concepts are clear.** Running a large survey to test relationships among concepts that have not been clearly defined produces precise numbers whose meaning is uncertain. Theory-building work, often qualitative, should come first.

**Endless building without testing.** Fields can accumulate frameworks and typologies that are never tested against each other. Without testing, there is no way to know which frameworks are useful.

## Implications for applied research

Commercial research often moves between building and testing without labelling either. Exploratory interviews about why customers switch suppliers are theory building: they develop an explanation of switching. A subsequent survey that measures the proposed drivers and models their effect on switching is theory testing. Being explicit about which stage a study is at helps set expectations. An exploratory study should not be expected to estimate the size of effects, and a survey should not be expected to discover factors it did not measure. Planning research as a sequence, with each stage designed for its own purpose, is usually more efficient than asking a single study to do everything.

## References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496–515. https://doi.org/10.5465/amr.1989.4308374

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal, 50*(6), 1281–1303. https://doi.org/10.5465/amj.2007.28165855

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review, 36*(1), 12–32. https://doi.org/10.5465/amr.2009.0486

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review, 32*(4), 1246–1264. https://doi.org/10.5465/amr.2007.26586086

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*(4), 806–834. https://doi.org/10.1037/0022-006X.46.4.806

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

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly, 40*(3), 371–384. https://doi.org/10.2307/2393788

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory, 30*(3), 167–186. https://doi.org/10.1177/0735275112457914

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review, 14*(4), 516–531. https://doi.org/10.5465/amr.1989.4308376

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371