

SURVEY DEVELOPMENT

A Theory-Driven Mixed-Method Approach

**TONY CHIU MING LAM AND
KATHY ELLEN GREEN**



“This book is carefully written to ensure precise meaning and that it is understandable to a broad audience. In my more than three decades following survey research, no other text has addressed the topic as intimately. Although there are many books on survey research, they lack the organizing schema of being theory based. While it is easy to tell what things are or how to do them, without the basis in theory (and empirical evidence) the reader is left without the framework to understand why they are that way. Developing this understanding is what equips your audience to move beyond the specifics that are presented to address new and different challenges as they go about designing and conducting survey research and making sense of their results. This moves this book to the forefront of the many texts available.”

Steve Siera, *Ph.D., former Chair, AERA SIG on Survey Research in Education, former Interim Dean and Associate Dean College of Education and Counseling Psychology, St. Martin's University, Lacey, Washington, USA*

“Required reading in multi-disciplinary seminars in survey methodology and research design, the proposed Theory-Driven Mixed Method (TDMM) approach is a pragmatic, timely, and useful contribution in a fast-transforming field. Through compelling content, eclectic examples, brilliantly designed illustrations, and pithy tables, the authors draw upon their extensive experiences and passion for the subject matter. This must-have tool in researchers' armory is a critical reminder for the need to balance survey design features and survey quality.”

Kranti (Kran) Dugar, *Ph.D., Assistant Professor of Marketing, College of Business, University of Wisconsin-Eau Claire, USA*



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Survey Development

Survey Development: A Theory-Driven Mixed-Method Approach provides both an overview of standard methods and tools for developing and validating surveys and a conceptual basis for survey development that advocates establishing and testing of hypotheses pertaining to presumptions and score-interpretation and use inferences and mixing quantitative and qualitative methods.

The book has 14 chapters which are divided into four parts. Part A includes six chapters that deal with theory and methodology. Part B has five chapters and it gets into the process of constructing the survey using both quantitative and qualitative methods. Part C comprises two chapters devoted to assessing the quality or psychometric properties (reliability and validity) of survey responses. Finally, the one chapter in Part D is an attempt to present a synopsis of what was covered in the previous chapters in regard to developing a survey with the TDMM framework for developing survey and conducting survey research. This provides a full process for survey development intended to yield results that can support valid interpretation and use of scores.

Including detailed online resources, this book is suitable for graduate students who use or are responsible for interpretation of survey research and survey data as well as survey methodologists and practitioners who use surveys in their field.

Tony Chiu Ming Lam is Associate Professor of Applied Measurement and Program Evaluation at the Ontario Institute for Studies in Education, University of Toronto, Canada

Kathy Ellen Green is Professor Emeritus, Research Methods and Statistics, Morgridge College of Education, University of Denver, USA



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Survey Development

A Theory-Driven Mixed- Method Approach

Tony Chiu Ming Lam and Kathy Ellen Green

First published 2023
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2023 Tony Chiu Ming Lam and Kathy Ellen Green

The right of Tony Chiu Ming Lam and Kathy Ellen Green to be identified as authors of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

IBM SPSS Statistics software (“SPSS”) is a registered trademark of IBM Corp. Screen images copyright © IBM Corp. Reprint Courtesy of IBM Corporation©. This book is not approved or sponsored by SPSS Statistics.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-367-22232-1 (hbk)

ISBN: 978-0-367-22233-8 (pbk)

ISBN: 978-0-429-27392-6 (ebk)

DOI: 10.4324/9780429273926

Typeset in Minion
by codeMantra

Access the Support Material: www.routledge.com/9780367222338

Contents

	Preface	xiii
	Acknowledgments	xvii
Part A	Theory and Methodology	1
1	Survey Research	3
	1.1 Overview	3
	1.2 Research	3
	1.2.1 Variables in Research	4
	1.2.2 Experimental Designs and Causal Interpretation	11
	1.3 Survey	12
	1.3.1 Scale	14
	1.3.2 Index	17
	1.4 Survey Research	19
	1.4.1 Survey Research Design	20
	1.4.2 Survey Research with Scales	23
	1.4.3 Ethics Concerns in Survey Research	24
	1.5 Chapter Summary	26
2	Theory and Validity	27
	2.1 Overview	27
	2.2 Theorizing in Measurement Validation and Application to Survey Development	28
	2.3 The Role of Theory in the Evolution of the Concept of Validity	30
	2.3.1 Validity Trinity (Atheoretical Framework)	31
	2.3.2 Validity Duality (Construct Validity Centered)	32

	2.3.3	Unitary Validity (Argument- and Interpretation-Centered)	40	
	2.4	Chapter Summary	42	
3		Mixed Methods		44
	3.1	Overview	44	
	3.2	Mixed Methods in Survey Research	45	
	3.2.1	Mixed Methods Designs in Survey Research	47	
	3.2.2	Three Examples of Survey Development	50	
	3.2.3	Synthesizing Qualitative and Quantitative Data	52	
	3.2.4	Recap	53	
	3.3	Chapter Summary	55	
4		Measurement Theory		56
	4.1	Overview	56	
	4.2	A Brief History of Testing and Survey Research	58	
	4.3	True-Score/Classical Test Theory (CTT)	60	
	4.4	Item Response Theory (IRT)	61	
	4.4.1	Major Characteristics of IRT	61	
	4.4.2	Estimation of Logits	67	
	4.4.3	Identification of Misfits	70	
	4.4.4	Assumptions of IRT Models	71	
	4.5	Comparison between IRT and CTT	72	
	4.6	Chapter Summary	73	
5		Theories of Response and Error		74
	5.1	Overview	74	
	5.1.1	Random and Systematic Errors	76	
	5.1.2	Sampling and Response in Survey	78	
	5.1.3	Design and Implementation in Survey	78	
	5.1.4	Theories of Response and Error	80	
	5.2	A Theoretical Framework of Sources of Survey Errors	82	
	5.3	Theories of <i>Why</i> People Respond to Surveys and <i>Why</i> Sampling Errors Occur	83	
	5.3.1	Theories of Why People Respond	83	
	5.3.2	Theories of Why Sampling Errors Occur	87	
	5.3.3	Implications of Theories of <i>Why</i> People Respond and <i>Why</i> Errors Occur for Minimizing Sampling Error and Maximizing Response Rates	89	
	5.4	Theories of <i>How</i> People Respond and <i>How</i> Response Errors Occur	92	
	5.4.1	Instrument as the Source of Response Design Error (RDE)	93	
	5.4.2	Participant as a Source of Response Implementation Error (RIE)	97	

	5.4.3	Context as a Source of Response Implementation Error (RIE)	111
	5.5	Chapter Summary	114
6		Participant Sampling	116
	6.1	Overview	116
		6.1.1 Selection and Matching	117
		6.1.2 Focus of this Chapter	119
	6.2	Whom Should We Select?	119
		6.2.1 Define the Population	119
		6.2.2 Breadth of the Population	120
		6.2.3 Describe the Population	122
	6.3	How Should We Select?	122
		6.3.1 Probability Sampling	124
		6.3.2 Non-Probability Sampling	130
	6.4	How Accurate and Precise Are the Sample Estimates?	132
		6.4.1 Sample Statistics	133
		6.4.2 Precision of Sample Statistics	134
	6.5	How Many Should We Select?	138
	6.6	What Evidence Do We Provide to Demonstrate How Representative Our Sample is of the Population?	141
	6.7	Chapter Summary	142
Part B		Survey Construction	145
7		Survey Development Framework	147
	7.1	Overview	147
	7.2	Survey Blueprint (SB)	150
		7.2.1 Survey Development	150
		7.2.2 Survey Administration	156
	7.3	Interpretation/Use Argument (IUA)	157
		7.3.1 Claims Preparation	160
		7.3.2 Claims Justification	173
		7.3.3 IUA Evaluation	186
		7.3.4 Summary of Guidelines for Performing IUA-Based Approach to Validation	188
	7.4	Chapter Summary	190
8		Writing Items	193
	8.1	Overview	193
	8.2	Item Types	194
		8.2.1 Open-Ended Item	194
		8.2.2 Closed-Ended Items	197
		8.2.3 Sensitive Items	211

8.3	Recommendations and Resources for Generating Items	213
8.3.1	Sources of Existing Items	214
8.3.2	Creating New Items	215
8.4	Chapter Summary	219
9	Constructing Response Scales	220
9.1	Overview	220
9.2	Scaling Methods	220
9.2.1	Thurstone	221
9.2.2	Likert	222
9.2.3	Guttman	224
9.2.4	Semantic Differential	225
9.3	Issues in Choosing the Response Scale for Rating Items	226
9.3.1	What Type of Anchors Should Be Used?	226
9.3.2	How Many Response Options or Scale Points Should Be Used?	227
9.3.3	What Specific Anchors of the Chosen Type Should Be Used?	229
9.3.4	In What Order Should Verbal Anchors Be Presented?	235
9.4	Chapter Summary and Recommendations	236
10	Qualitative Survey Item Review	239
10.1	Overview	239
10.2	Qualitative Item Review Methods	240
10.2.1	Ex Ante Method	241
10.2.2	Individual-Based Qualitative Item Review Methods	241
10.2.3	Group-Based Qualitative Item Review Methods	248
10.3	Survey Format and Survey Review	252
10.3.1	Survey Format	253
10.3.2	Item Presentation Order	253
10.3.3	Survey Review	254
10.4	Chapter Summary	255
11	Quantitative Survey Item Review	256
11.1	Overview	256
11.2	Quantitative Item Review Methods	256
11.2.1	Ex ante Methods	257
11.2.2	Individual-Based Quantitative Item Review Methods	259
11.2.3	Group-Based Quantitative Item Review Methods	264

11.3	Comparison of Qualitative and Quantitative Item Review Analysis Methods	286
11.4	Mixing Quantitative and Qualitative Methods in Item Review Analysis	292
11.4.1	Recommended Path for Item Review	293
11.5	Chapter Summary	294
Part C	Assessing Psychometric Properties	295
12	Assessing Reliability	297
12.1	Overview	297
12.2	Reliability of Scale Scores	298
12.2.1	Internal Consistency Reliability	299
12.2.2	Test-Retest Reliability	303
12.2.3	Equivalent-Forms Reliability with and without Time Lapse	304
12.3	Analysis of Consistency of Scores	306
12.3.1	Consistency of Item Responses to Categorical Variables	307
12.4	Inter-Rater Reliability	307
12.4.1	Categorical Items	307
12.4.2	Interval Scale	309
12.4.3	Factors Influencing Inter-Rater Reliability	311
12.5	Chapter Summary	311
13	Assessing Validity	313
13.1	Overview	313
13.2	Idiographic Validity Evidence	315
13.2.1	Content Evidence	315
13.2.2	Factor and Cluster Structure Evidence	320
13.3	Nomological Validity Evidence	326
13.3.1	Association Evidence	327
13.3.2	Causal Modeling Evidence	330
13.4	Chapter Summary	336
Part D	Recommendations for Theory-Driven Mixed-Methods Survey Development and Validation	337
14	Synopsis	339
14.1	Overview	339
14.2	Criteria for Developing the TDMM Approach to Survey Development	339
14.2.1	Careful Proactive Thinking and Action	340
14.2.2	Guided by Theory	340

14.2.3	Distinguish between Idiographic Validity and Nomological Validity	340
14.2.4	Mix Quantitative and Qualitative Methods	340
14.2.5	Emphasize the Need to Identify and Minimize All Potential Errors	341
14.3	Using Theory to Construct Survey and Validate Responses	341
14.3.1	Components of the TDMM Survey Development Framework	342
14.3.2	Procedure for Implementing the TDMM Survey Development Framework	344
14.4	Mixing Qualitative and Quantitative Methods within a Theoretical Structure in Survey Development	345
14.4.1	Survey Design	345
14.4.2	Survey Construction	351
14.4.3	Pilot Testing	351
14.5	Example of a Hypothetical Survey Research Study with the TDMM Approach	356
14.5.1	Survey Development	356
14.5.2	Survey Administration	367
14.6	Chapter Summary	368
	References	371
	Author Index	393
	Subject Index	398

Preface

Conducting surveys is fascinating, seems easy, gets results fairly quickly, and so is a very common data collection method chosen by education, health, business, and social science researchers in performing their work. We have been privileged to support a number of master's and doctoral students as methodologists on the students' committees, and our experiences in doing so shaped the inspiration for this book. Additionally, we were motivated to capture and advocate for the latest transformations in survey research and development methodology that had not been included elsewhere.

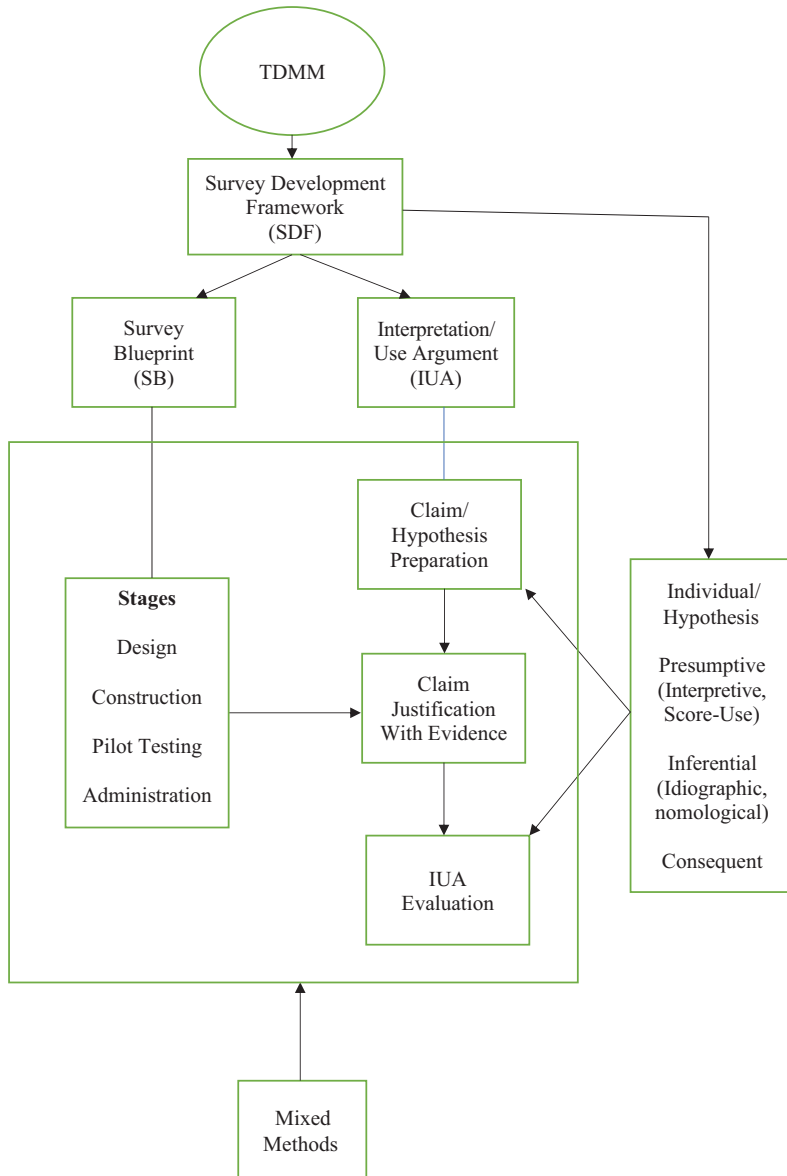
The two pillars or themes of our book are: (1) *utilizing theory* and (2) *intermixing quantitative and qualitative methods*. We believe that theorizing forces researchers to delve proactively and deeply into the issues under investigation, and mixing methods expands the repertoire of tools for researchers to choose from to collect data and validate responses. These two themes form the foundation of our book and they are reflected in the title: *Survey Development: A Theory-Driven Mixed-Method (TDMM) Approach*. Other features include requiring researchers to proactively design and anticipate findings and problems upfront, implementing the interpretation/use-argument validation method with consideration of both idiographic and nomological validities, and attending to and minimizing all sources of errors. Below is a concept map showing the major components and structure of the TDMM approach to survey development:

Much has been published to guide researchers through the process of developing surveys and conducting survey research. What is typically missing in those sources is emphasis on theory to conceptualize and guide the research and on using both quantitative and qualitative methods to develop surveys, collect data, and interpret findings. The methods we prescribe primarily stem from the quantitative paradigm. Also, we hope that the title of our book does not imply that our book has the definitive word on this approach. Rather, it should be treated as an illustration of how surveys are constructed, validated, and implemented, relying on hypothesis testing and with both quantitative and qualitative data. In addition to this feature of our book, there are three more that distinguish it from other survey development books. First, missing in survey development texts is an explanation of how measurement theory and statistical methods are used in survey research, especially item response theory. Our book fills this void. Secondly, although our book title refers to survey development, we describe survey development within the context of survey research by including a chapter on survey research designs and a chapter on

participant sampling. Third and finally, we include theories of response and error and extensive qualitative and quantitative survey review methods that are not typically found in other survey development books.

Our book is divided into four parts. Part A includes six chapters that deal with theory and methodology: we describe and define survey research (Chapter 1), present the concept and the different types of theories and how theorization is fused into the validation methodology (Chapter 2), define mixed methods and the different designs of mixing methods (Chapter 3), briefly explain measurement theory and how it is used to account

Theory-Driven Mixed Methods (TDMM) Approach to Survey Development



for the origin of observed scores (Chapter 4), and present theories of how and why we respond and make errors (Chapter 5) and methods of sampling participants (Chapter 6). Part B gets into the process of constructing the survey, which Chapter includes preparing a survey development framework (Chapter 7), writing items (Chapter 8), constructing response scales (Chapter 9), qualitative survey review (Chapter 10), and quantitative review (Chapter 11). Part C is devoted to discussing the quality or psychometric properties of survey responses. The two chapters in this section are assessing reliability (Chapter 12) and assessing validity (Chapter 13). Finally, Part D (Chapter 14) is our attempt to present a synopsis of what we covered in the previous chapters in regard to developing a survey with the TDMM approach or framework. Below, we present a concept map of the major components of the framework.

We have both taught a course in survey research methods for a number of years and have done research on surveys, which was a good jumping off place for writing this book. Kathy has taught measurement and statistics while Tony has taught program evaluation and research, so our complementary backgrounds helped us write in those areas. We both are very much aware that we are only scratching the surface in terms of depth in some areas. Readers may note the “beyond the scope of this book” comments here and there. We intended the book to be accessible to a general audience, though having knowledge of measurement and statistics is always helpful. Our book can be used as a text for survey research methods course or as a reference for those engaged in survey research or use surveys for their research. The supplemental material provided online are designed to help in conducting some of the recommended analyses.

This is the first time anyone has attempted to develop an approach to survey development that is theory driven with mixed methods. The framework we proposed is conceptually straightforward but gets complicated with plenty of nuances in operation and we are certain that there is much room for improvement. We hope that in a small way our book will encourage you to carry on what we and others have started.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Acknowledgments

Tony: I am grateful for the opportunity to present in this book a new approach to survey development and I am gratified knowing that my parents would be pleased and my wife Leslie would be relieved that, despite obstacles, the book is now finished.

Kathy: My thanks and appreciation go to members of the AERA Special Interest Group: Survey Research who invited me to join the world of survey research. Thanks to my students through the years who asked intriguing questions and have helped me learn so much. Thanks to family and friends who commiserated with me and who are no doubt happy the book-writing process is complete.

And, thanks to our patient and supportive editor Hannah Shakespeare who is a master at quick response to emails, clear answers to questions, and helped us through the process from the beginning.

Finally, this book contains multiple images from IBM® SPSS® Statistics Software (SPSS). These images are reprinted courtesy of International Business Machines Corporation, © International Business Machines Corporation. Permission to copy Winsteps output and documentation is granted by Winsteps.com, Copyright © 2022 by John M. Linacre.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part A

Theory and Methodology



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

1

Survey Research

1.1 OVERVIEW

This book is about developing surveys for survey research and for all other types of research that use surveys to collect data. It seems pertinent that we begin the book defining and explaining survey research and survey in this first chapter. With that in mind, we open this book with this question: *what is survey research?* To answer this question, we need to understand what research is and how survey research differs from the other types of research. To accomplish this goal, we do the following in this chapter:

- Define research.
- Differentiate between antecedent, mediating, consequent, and moderating variables and how they are used to define experimental and non-experimental research.
- Explain what survey, scale, and index are.
- Define survey research.
- Describe various survey research designs.
- Consider ethics concerns unique to survey research.

1.2 RESEARCH

Persistence is implied if we search and re-search for something. That's what the term “**research**” signifies, a consistent effort to find something—doing research requires consistent effort! We want to preface our book with this fact, and we hope that you, our readers, take this to heart before beginning your journey with us.

As used to denote a field of activities in which people engage, research entails finding answers to questions that arise out of either curiosity (e.g., why do cats meow? What is the meaning of life?) and/or necessity (e.g., how do we control our environment and climate? How do we educate and prepare young people to function effectively in society?). To enhance efficiency and increase probability of locating answers that are correct, we apply scientific canons and compile what we found from repeated searches (or re-search). Against this backdrop, simply put, “*research*” is a *discipline of coordinated, systematic, and*

scientific attempts to gather information to respond to meaningful questions in an effort to advance our common knowledge or to solve problems.

Research in social sciences, health, and business is divided into two groups by the world view to which researchers adhere: *qualitative research* and *quantitative research*. The paradigm that qualitative researchers uphold considers reality as something that is determined by how we interpret our experiences and, hence, there are as many realities as there are perspectives. However, the paradigm quantitative researchers abide by endorses an objective reality, and hence, there is only one perspective that is correct. This paradigmatic divide has forced researchers to choose and to function within the group they prefer with the methods and procedures stipulated by that group. Recognizing the pragmatic disadvantage of confining researchers to methodologies endorsed by only one group, the mixed research methods campaign started decades ago broke down the paradigmatic barriers to create a venue in which both qualitative and quantitative methods are allowed and in fact encouraged. As quantitative researchers, we (both authors of this book) chose this venue to adopt qualitative methods in our framework of constructing and validating surveys. We believe that the confluence of methodologies from and insights generated by both method types offer us a better chance of creating sound surveys.

Research that involves people as participants has two major components: measurement and design. By *measurement*, we mean using a variety of methods and procedures such as observation, testing, and surveying, to collect data and to ensure that those data are reliable and valid. By *design*, we mean the process of recruiting and assigning participants to perform different tasks at different times (such as before and after a treatment), with different frequencies (such as the number of times different research studies engage participants in post-treatment assessment of their change in performance), and under different contexts (such as the difference in treatment received by the experimental and control participants, or a research study being conducted in urban or rural settings). Mixing methods implies using both qualitative and quantitative methods to collect data and to design research. For developing surveys, the focus centers on mixing measurement strategies rather than designs. In Chapter 3, we expand on our discussion regarding how qualitative and quantitative (measurement) methods work jointly to develop surveys.

1.2.1 Variables in Research

The essential task of a quantitative researcher is to describe the world objectively, often by examining the causal or non-causal relationships (or differences) among some selected variables. In this way, you can say that research is a deliberate act of studying variables, measuring variables, and analyzing relationships among variables. So, what is a variable as opposed to a constant?

A *variable* refers to an attribute (e.g., age) whose levels or values change and they change in two different ways. First, change can occur within an entity, which we refer to as *intra-individual* or *within-individual* or *intra-subject variability*, e.g., age changes for an individual across time as they grows older. Second, change can occur between entities, which we refer to as *inter-individual* or *between-individual* or *between-subject variability*, e.g., in a group of individuals, the ages differ across the group members. (See a discussion below for an elaboration of attribute.) The same distinction (intra-, inter-) can be applied to other variables like math achievement, self-efficacy, and so on. However, a *constant* is a state of affairs that does not vary, e.g., basketball player number, a name, or a tattoo.

1.2.1.1 The Four Types of Variables

In any research study, four categories of variables are explicitly or implicitly involved. They are **antecedent, mediating, moderating, and consequent variables**. These four types of variables are depicted in a generic causal model in Figure 1.1.

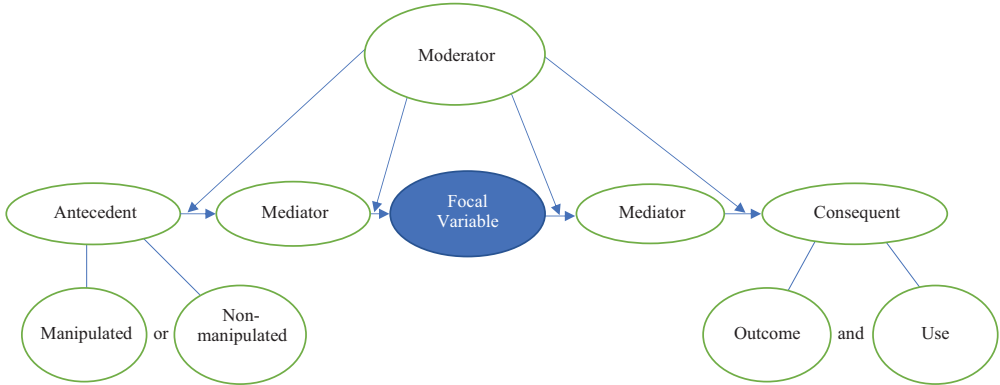


Figure 1.1 A Generic Causal Theory Depicting the Different Types of Variables in Research.

From Figure 1.1, we can make the following observations:

- *Antecedent* variables could be manipulated (experimental) **or** not manipulated (non-experimental) and we are interested in their effects on other variables (antecedent variable →). In quantitative research, an antecedent variable is referred to as an *independent* or *treatment* variable.
- *Consequent* variables refer to both (1) outcome variables that are affected by other variables of interest to us (→consequent variable), e.g., academic achievement as affected by class learning activities or homework assignments) and (2) the consequence of using the results from a research study (e.g., funding schools based on student performance on a standardized test) or from measurement (e.g., measuring teachers' self-efficacy and labeling some teachers as having low self-efficacy and as a consequence demoralizing them). In quantitative research, a consequent variable is referred to as a *dependent* or *outcome* variable.
- *Moderating* variables, like antecedent variables, are variables that affect other variables but they are not affected by other variables in a study (moderating variable→). For example, age or prior learning affects students' motivation to learn.
- We refer to both moderating and antecedent variables as *exogeneous* variables (variables in a causal map that only affect other variables).
- *Mediating* variables are variables that affect other variables and also are affected by other variables (→mediating variable→). In quantitative research, a mediating variable is also referred to as an *intervening* or *intermediate* variable. For example, attitude is affected by training and change in attitude affects learning.
- We refer to consequent and mediating variables as *endogenous* variables, i.e., variables in a causal map that are affected by other variables. The difference between these two types of variables is that consequent variables are variables that are only affected by other variables, whereas mediating variables are both affected by and affect other variables.

A variable can serve as one type of variable in one study but another type in a different study. For example, “student teacher-interaction” may serve as a mediating variable in one study,

Classroom management style → student-teacher-interaction → student learning,
and as a consequent variable in another

Teacher training → classroom management skill → student-teacher-interaction

Among these four types of variables, the distinction between mediating variable and moderating variable is murky to a lot of students we encountered. Let us elaborate on these two types of variables next and add a few more concepts at the same time.

1.2.1.2 Mediating Variable

Mediating variables connect two variables in a causal chain and they help us understand the process through which two variables are related. For example, suppose there is an association between sex and majoring in math in college students. In this non-experimental causal model, the effect of “sex” on “majoring in math” may be mediated by “attitude towards math” as shown in this causal chain:

Sex → Attitude towards math → Majoring in math

In this case, sex affects attitude and attitude affects majoring in math. These two effects are *direct* as there are no other variables between the two variables. However, “sex” has an *indirect* effect on “majoring” because the “attitude” variable is between these variables. To interpret, “sex” affects “majoring” by affecting “attitude.” In this example, the “attitude” variable is the mediating variable in this causal chain, “sex” is the antecedent variable, and “majoring in math” is the consequent variable. So, one can consider a mediating variable as a special type of consequent variable since it is affected by an antecedent variable. In our example, the mediating variable “attitude” is affected by “sex.”

A mediating variable can sometimes fully account for the relationship between an antecedent and consequent variable. In our example, sex has an *indirect* effect on majoring in math, and attitude has a *direct* effect on majoring in math. If we control the attitude variable (by converting the variable statistically into a constant), the correlation between sex and majoring in math may disappear.

Here is another example of mediating variables. Suppose a teacher training workshop on student-centered instruction affects teacher’s attitude, knowledge, and, subsequently, teaching performance in that order.

Workshop → Attitude → Knowledge → Performance

In this example, “attitude” and “knowledge” are the two mediating variables between “workshop” and “performance”; “attitude” is the mediating variable between “workshop” and “knowledge”;

Workshop → Attitude → Knowledge

“knowledge” is the mediating variable between “attitude” and “performance.”

Attitude → Knowledge → Performance

Both “workshop” and “attitude” have *indirect* effects on “performance” (“workshop” affects “performance” through its effect on “attitude” and “knowledge”; “attitude” affects “performance” by affecting “knowledge”), and “knowledge” has a *direct* effect on performance. If we statistically hold both “attitude” and “knowledge” variables constant, the effect of workshop on performance may diminish.

As mentioned, mediating variables explain the underlying mechanism that connects two variables. Since understanding why and how an effect occurs is the crux of conjecturing about theory, identifying mediating variables is a *requirement* in theory-driven survey development and research.

1.2.1.3 Moderating Variables

Moderating variables can affect the strength and direction of the linkage between two variables such as an antecedent variable and a consequent variable. For example, referring to the above two examples, the impact of “sex” on “majoring” may only hold for “high academic ability” students, and the causal impact of “workshop” on “performance” may only hold for females but not for males (which would be seen as an interaction in a statistical model). The ability variable in the first example and the sex variable in the second example are moderators. (Note: sex is the antecedent variable in the first example.)

Not only do moderating variables serve as marker variables for describing the population (to be discussed in Chapter 6 on Sampling), examination of moderating variables has implications for identifying sources of error in measurement and for identifying sources of variability for isolating treatment impact on outcomes and for knowledge generation. We will discuss these points about moderating variables next.

When a moderating variable is a source of measurement error, the relationship between the moderating variable and a focal variable is an artifact due to measurement error. However, it is a source of variability of results if the relationship is real. For example, if age is found to be related to or has an impact on emotional intelligence and we conclude that older respondents’ emotional intelligence is higher than that of the younger respondents. Age is a *source of measurement error* if the observation is due to satisficing (invalid responses) by the younger respondents. Age is a *source of variability* if both groups of respondents are attentive and there is indeed a true difference in emotional intelligence between the two groups. We need to determine if a moderating effect is an artifact or real in a study. If it is due to error, we want to eliminate its effects from results of the study. Apart from being a source of measurement error, moderating variables could also be a source of error in interpretation.

To determine the effect of an antecedent variable with or without manipulation, influences of all variables except the treatment variable must be removed or minimized, otherwise it becomes a source of error in interpretation. For example, suppose the antecedent variable is “classroom management skill” as measured by teachers’ self-assessment of learning from a training workshop. If we suspect that “teacher experience” is a moderating variable that can affect learning, we would implement some procedure to control the effect of teacher experience (such as measuring it and removing its effect statistically with a statistical procedure like analysis of covariance). In this situation, the teacher experience moderating variable is a source of interpretation error. Sources of error (measurement and interpretation) are referred to by these multiple terms: *extraneous*, *nuisance*, *control*, *concomitant*, *covariate*, and *confounding* variable. As you will learn in Chapter 7, we need to include effects of potential extraneous variables in a study as presumptions to be tested with evidence and eliminate these effects if found.

However, we may want to study the differential effect of teacher experience on learning for future training planning purposes. In this case, we want to deliberately include the teacher experience variable in the research design as an antecedent variable (a source of variability) and to find out if teacher experience is correlated with learning; in other words, if teachers with different levels of experience learned differently, or, if findings from teachers at one level of teacher (e.g., first year teacher) can be *generalized* to teachers at other levels (e.g., teachers who have been teaching for ten years).

With these two ways of managing the variable in experimental research, effects of a moderating variable (as a source of variability) are either avoided (by considering it as an extraneous variable) or are the subject of study (by considering it as an antecedent variable). The latter method of handling a moderating variable is preferred (if a study can handle a multitude of variables) because it is informative. Let us explain.

The two primary purposes that drive research are for practical needs with policy decision implications and for knowledge gain with theoretical implications. Findings from moderating variable research for practical use could inform needed research for theoretical advancement, and *vice versa*. For example, if a research study is interested in finding out if a medication works and if it works equally well for men and women for drug approval and prescription purposes, we need to collect data from both men and women even if we believe there to be no theoretical relationship of sex to the outcome. However, if we find the medicine benefits females but not males (i.e., a treatment by sex interaction is noted), findings from this research may pique the interest of researchers to investigate why the findings are observed.

In another situation, to study determinants of learning, class size has been found to be correlated with students' academic achievement so we should know the size of classes students are in for our theoretical understanding. This theoretical investigation could have implications for policy decisions. If a positive finding with a small class size does not generalize to a large class size, policy may be proposed to budget hiring of teachers in an attempt to keep class size relatively small. As seen in these two examples, consideration of moderating variables is critical in designing research in the field environment in which a multitude of variables are operating outside of the researchers' control that can affect the findings.

1.2.1.4 *Comment and Summary*

A few points to highlight:

- A variable can perform a different function in different research studies. For example, in the above examples, sex is the antecedent variable in one research study and a moderating variable in another.
- All but consequent variables can be manipulated or controlled.
- Moderating variables could be sources of variability, sources of measurement error, or sources of interpretation error.
- Not all studies include all four types of variables, although we strongly encourage that all types be considered.
- A study may include more than one of a certain type of variable, i.e., multiple antecedent, mediating, consequent, or moderating variables.

To close this section, we summarize the definitions and provide some examples of the four types variables in Table 1.1.

Table 1.1 Variable Definition and Examples

Variable	Definition	Example
antecedent (Independent, (Independent treatment))	Variables <i>created and controlled</i> by researchers whose <i>effects on</i> other variables the researcher wants to assess.	<ul style="list-style-type: none"> • Early education programs • Teacher in-service training programs • A new depression medicine • Training programs aimed to induce teachers' empathy or self-efficacy. <p>For example, in a treatment-control group design, <i>field work experiences of preservice teachers</i> in the experimental group are <i>manipulated</i> to increase efficacy beliefs in mathematics, whereas such experience of preservice teachers in the control group are not manipulated. After the practicum, math self-efficacy of the preservice teachers in both groups is measured as a manipulation check. And after one year of teaching, math performances of students (consequent variable) in the classes of experimental and control teachers are compared.</p>
Non-Manipulated	Variables <i>not created and controlled</i> by researchers whose <i>effects on</i> other variables the researcher wants to assess	<ul style="list-style-type: none"> • Sex effect • Effects of different <i>types of schools</i> (public versus private) • Effects of <i>students' existing level of motivation to learn</i> on achievement • Effects of teachers existing <i>empathy</i> levels on student learning.
Mediating (Intervening, intermediate)	Variables that are <i>affected by</i> antecedent variables and <i>affect</i> consequent variables	<p>With manipulated antecedent variables:</p> <ul style="list-style-type: none"> • Early education programs stimulate childrens' <i>curiosity and on-task behavior</i> which, in turn, affects language proficiency acquisition. • Teacher training programs enhance <i>teachers' attitude, efficacy, knowledge, and skills</i> which, in turn, affect their ability to use student-centered instruction. • A new anti-depression medicine improves <i>patients' confidence about recovery</i> which, in turn, affects depression. • An in-service teacher training program in science boasted participant teachers' <i>self-efficacy in science</i>, which enhanced these teachers' teaching performance. <p>With non-manipulated antecedent variables:</p> <ul style="list-style-type: none"> • Sex affects <i>academic interest</i> which, in turn, affects choice of majors in college;

(Continued)

Table 1.1 (Continued)

Variable	Definition	Example
<p>Moderating (as source of interpretation error) (Concomitant, covariate, confounding, control, extraneous)</p>	<p>Variables that are <i>unaffected</i> by antecedent variables but that affect mediating and/or consequent variables.</p> <p>Quite often they are considered as nuisance variables whose effects in a research study researchers want to minimize. It is because their effects can confound the research findings. For example, the effect of an in-service teacher training program could be affected or confounded by teachers' teaching experiences or sex.</p>	<ul style="list-style-type: none"> • School type affects <i>attitude toward learning</i> which, in turn, affects students' learning; • Students' existing motivation to learn affects <i>students' on-task behavior</i> which, in turn, affects student academic performance. • Teacher empathy levels affects <i>student-teacher interaction</i> which affect student learning <p>With Manipulated antecedent variables:</p> <ul style="list-style-type: none"> • In an early education program designed to promote language proficiency, the program effect is moderated or affected by <i>parental involvement</i>. • The effects of a teacher in-service training program on teachers' ability to implement a student-centered instruction is moderated by <i>teachers' teaching experience</i> and sex. • The effects of an anti-depression medicine on depression are moderated by <i>patients' health status and desire for improvement</i>.
<p>Consequent Dependent (outcome)</p> <p>Results from using findings</p>	<p>With Non-Manipulated antecedent variables:</p> <ul style="list-style-type: none"> • Aptitude's effect on choice of majors in college is moderated by <i>sex</i>. • Students' learning as affected by school type is moderated by students' <i>socio-economic status</i>. • The effects of motivation to learn on academic performance are moderated by students' <i>age</i> and <i>ethnicity</i>. • The correlation between years of teaching experience and student academic performance is moderated by teacher <i>self-efficacy</i> and <i>empathy</i>. <p>An in-service teacher training program is designed to enhance teachers' empathy. Before and after training, a scale is administered to measure participating teachers' empathy levels at those two time points. The change in <i>empathy</i> is used as the outcome of the training program.</p> <p>Knowing and being labeled as having high or low empathy affects teachers' performance in the classroom.</p>	<p>With Non-Manipulated antecedent variables:</p> <ul style="list-style-type: none"> • Aptitude's effect on choice of majors in college is moderated by <i>sex</i>. • Students' learning as affected by school type is moderated by students' <i>socio-economic status</i>. • The effects of motivation to learn on academic performance are moderated by students' <i>age</i> and <i>ethnicity</i>. • The correlation between years of teaching experience and student academic performance is moderated by teacher <i>self-efficacy</i> and <i>empathy</i>. <p>An in-service teacher training program is designed to enhance teachers' empathy. Before and after training, a scale is administered to measure participating teachers' empathy levels at those two time points. The change in <i>empathy</i> is used as the outcome of the training program.</p> <p>Knowing and being labeled as having high or low empathy affects teachers' performance in the classroom.</p>

1.2.2 Experimental Designs and Causal Interpretation

In quantitative research, *experimental research* (including program evaluation) involves a deliberate and planned intervention (or treatment) that is *manipulated* by the researchers (e.g., a teacher training program on student-centered instruction). Referring to Figure 1.1, the antecedent (independent) variable is manipulated in experimental research and its causal effect on the consequent (dependent) variable is examined. The extent to which independent variables affect change in dependent variables is a concern of *internal validity* (see Shadish et al., 2002). The cornerstone of impact assessment is provision of a *counterfactual*, which is an estimate of the participants' performance or responses had they not received treatment. A counterfactual can be obtained through quantitative as well as qualitative methods (see Reichardt, 2022), and, if properly estimated, can eliminate threats to internal validity.

In quantitative research, manipulation of the treatment in an experimental design involves controlling what, when, and how the treatment is administered to whom (i.e., assigning recruited participants to different treatment conditions). When a control group is included in the experimental design and if the participants are randomly assigned to the treatment or control groups, the experimental design is referred to as a *true* or *strong* experimental design. If a control group is not formed by random assignment or if other strategies are used to estimate control group performance (also the counterfactual estimate), the experimental design is referred to as a *quasi*-experimental design. Finally, if an experimental design doesn't involve estimating performance of a control group, like the pre- and post-test single experimental group design, it is referred to as a *weak* or *pre*-experimental design.

Although non-experimental designs in which the independent variables are *not manipulated* are not set up to investigate causality like experimental designs with the assessment of a counterfactual, data collected from non-experimental research may be based on a theory that causally links different types of variables. If models are available that depict the causal impact of non-manipulated antecedent variables on mediating and consequent variables, and if all the variables are measured and the data are analyzed with special statistical analysis methods such that the results can reflect causal relationships among the variables involved, causal relationships can be determined with non-experimental data, even though the data are correlational in nature. The special statistical methods are path analysis and structural equation modeling (which are briefly described in Chapter 13).

The research mentioned above that examines the effect of sex on choice of math as a major is an example of non-experimental research, and more examples are provided in Chapter 2. *Nomological validity* is concerned with the testing of theory of correlations among multiple variables and the theory of causality which can be based on experimental as well as non-experimental data. (Nomological validity is discussed in the next chapter.)

In addition, a phenomenon can be studied with an experimental or a non-experimental research design. For example, we can study the relationship between meditation and mindfulness by surveying a group of people and finding out the correlation between their extent of practicing meditation and degree of mindfulness. Or, we can assign some people to receive meditation training and some not and compare their mindfulness after training. The former research is non-experimental and the latter experimental.

Regardless of the research design, a theoretical approach requires consideration of the structural relationship among the antecedent, mediating, moderating, and consequent variables. When a survey is used as the primary data collection method in non-experimental research, such research is considered survey research, which we will discuss after learning about what a survey is, which is our next topic.

1.3 SURVEY

In a social science or business research study, researchers collect data by reviewing existing documents, by observing people's behavior or interactions in natural and contrived environments, by administering tests, or by asking questions through surveys. A *survey* is a compilation of items assembled in a questionnaire that is designed to gather self-report data in either experimental or non-experimental research. Survey is basically a quantitative data collection method, even though it can be and almost always is supplemented by qualitative data. To analyze and understand survey as a data collection tool, we propose the following ten key characteristics or dimensions:

1. Survey data are *self-report data* about three different *targets* that are grouped in two categories:
 - *Respondent-Centered*: (1) Respondents report something about *themselves*, e.g., respondents' employment status or thinking process or attitude toward abortion.
 - *Stimulus-Centered*: (2) Respondents report about *others*, e.g., student evaluating teacher; (3) Respondents report something about some *objects* (e.g., a painting or trainer) or events (e.g., work environment or a concert).
2. Data collected by surveys could be *factual* information (e.g., crime rate, income, unemployment rates, household composition, and use of computer) or reflect an *abstract* attribute (e.g., attitude to school or reform-based instructional practices, learning styles, acculturation, ego-drive, need for cognition, empathy, emotional intelligence, personality type, impulsivity, risk-taking, locus of control, and self-efficacy).
3. The abstract attributes (or constructs) being measured by surveys could be related to affect (attitude, propensity, disposition, or interest), cognition (knowledge, cognitive processes), or psychomotor attributes (behavior, skills) that are formed through time.

Attitude is a feeling toward something or other individuals that may be reflected in behavior. *Propensity, disposition, or interest* is the tendency to do things, like habits and hobbies. *Knowledge* is an attribute that could be measured with an objective correct or incorrect answer (e.g., cholera is transmitted by staring at another person for 15 seconds, T/F) or a subjective self-report item (e.g., I know a lot about cholera). *Cognitive processes* are internal activities in performing a mental task (e.g., thinking about how to solve a math problem). Similar to knowledge, *behavior* may be measured with objective items such as items used to observe on-task behavior (e.g., number of time students are distracted) or with subjective items soliciting self-report about one's on-task behavior (e.g., I devote a lot of time performing the XYZ task). *Skills* refer to physical performance capacities (e.g., ability to perform a sport). Similar to behavior, skills can be measured with objective and subjective items.
4. Survey items may be *closed-ended* (e.g., do you support opening schools in the summer? yes/no) or *open-ended* (e.g., what are the reasons...?). While open-ended items are all quite similar in form, there are many varieties of closed-ended items such as magnitude estimation, rank order, multiple options, semantic differential, and rating items. Item types are described in Chapter 8. In a mixed method design, items are used to collect both quantitative and qualitative data.
5. *Unit of analysis* of survey data could be (1) score of a single independently operative item like a polling item (e.g., do you support increasing taxes to support the art

museum?), (2) summated or scale score, which is the sum or average of multiple item scores across all items in a “scale,” or (3) subconstruct score, which is the sum or average of multiple item scores across items in a subscale embedded in a larger scale.

6. Surveys can be conducted with different *modes of administration* which include (1) paper/pencil or computer-based questionnaire, (2) individual or group-based, (3) self-administered or administered by an observer, and (4) on-line or telephone or face-to-face survey. [The focus of this book is self-administered surveys though the information presented applies to interviews as well.]
7. The *research focus* or *statistics* generated from surveys could pertain to status or change in status, differences between groups, or the relationship between variables.

Statistics generated from surveys can be about (a) current *status* (e.g., percentage of respondents that support opening school for four-year-old children, average attitude toward student-centered instruction or teacher efficacy ratings), (b) *group differences* (e.g., difference in performance ratings between male and female teachers), or (c) *relationship* (correlation between teacher ratings and years of teaching), which could be correlational or causal.

Regarding the point above about the *research target*, we can say that surveys are designed to obtain statistics about the respondents, individuals other than the respondents, or external objects or events. Thus, survey research is generally quantitative in nature. Table 1.2 shows examples of research studies with different combinations of research foci and targets.

Table 1.2 Examples of Research Studies

Research Focus	Research Target		
	Respondent-centered		Stimulus-centered
	Respondents	Others	Objects
Status	Current employment rate in the respondents' population.	Student evaluation of teachers' performance with course evaluation questionnaire.	Ratings of quality of work environment.
Group Difference	Sex difference in employment rate in the respondents' population.	Difference in student evaluation of teachers' performance between private and public schools.	Difference in quality of work environment between private and government sectors.
Relationship ^a (Correlation)	Relationship between employment status and self-reported stress level in the respondents' population.	Relationship between student evaluation of teachers' performance & students' self-reported learning.	Relationship between quality of work environment and workers' self-reported job satisfaction.

^aPlease note that an analysis of relationship (e.g., relationship between sex variable and empathy variable) can be converted to an analysis of group differences (e.g., sex difference in empathy), and vice versa. Also, analysis of difference in the absence of treatment has also been referred to as causal comparative designs.

8. Surveys are generally used to collect data from a *sample* of a target population (whereas in a *census*, data are collected from all members of a population). Consequently, sampling correctly to assure that the sample of respondents is representative of the target population (or that findings based on a sample are generalizable to the population) is critical in survey research and in quantitative research in general.
9. *Frequency of administration* in surveys could either be *ongoing* (e.g., survey about pre-, post-, and six months later knee function with surgery) or on demand (e.g., public's opinion about immigrants).
10. As discussed above, the research context within which surveys are used could be experimental in which interventions and assignment of participants to different tasks are manipulated (e.g., to evaluate a professional development workshop) or non-manipulated (e.g., survey high school students' attitude toward climate control). Table 1.3 shows sample research questions about teachers' attitude toward student-centered instruction pertaining to the three research foci in experimental and non-experimental investigations.

Table 1.3 Examples of Research Questions

<i>Research Focus</i>	<i>Research Type</i>	
	<i>Experimental</i>	<i>Non-Experimental</i>
Status	How do teachers (participants) in the treatment and control group feel about student-centered instruction before and after the training program?	How do high school teachers feel about student-centered instruction?
Group Difference	What is the difference in attitude toward student-centered education between teachers in the treatment and control group?	What is the difference in attitude toward student-centered education between male and female high school teachers?
Relationship	What is the relationship between teachers' attitude toward student-centered education and level of participation in the student-centered-instruction training program?	What is the relationship between attitude toward student-centered education and teacher self-efficacy?

1.3.1 Scale

Any quantitative research study may use a survey comprising independent items and/or a group of related items designed to measure an attribute. This group of related items is referred to as a *scale*, and the attribute being measured is referred to as a *construct*. In this way, a survey may include (1) independent items alone, or (2) a scale and sometimes multiple scales, or (3) a combination of independent items and scales. Before discussing different types of research involving the use of scales, we first elaborate the concepts of scale and construct.

As discussed above, variables can be factual (e.g., marital status or income level) or *abstract* (e.g., opinion or judgment). The abstract variables are latent constructs and are not directly observable. These are concepts invented by humans to help us understand and communicate about phenomena around us. For example, attitude toward diversity reflects the degree to which we accept people as fellow citizens who are different from us in color of skin or gender or sexual preference, and this attitude varies across time and people.

To measure a construct (e.g., self-efficacy, anxiety, and attitudes toward climate control), we need to develop a *scale*, which is a self-report instrument that consists of (1) a set of closed-ended items that comprise both “statements” or “questions” and also (2) a “rating” or “response” scale that presents options for participants to choose from as their answers. (The focus of this book is about developing or using scales in surveys rather than about surveys comprising only standalone items.) As seen in this definition, a scale contains rating response scales. We should be careful not to refer to a scale as a rating scale—the scale comprises the items as well as a rating or response scale. However, scale has been commonly referred to as “survey/additive/summed/construct/measurement scale.”

A construct may comprise subconstructs, which are also not observable and which serve as components, factors, aspects, or facets of the construct. For example, “attitude towards school” might comprise opinions about (a) learning materials, (b) instructional methods, (c) learning facility, (d) quality of teachers, and (e) physical environment. These five subconstructs are the components of the construct “attitude towards school.” In this case, the scale to measure “attitude towards school” will comprise five subscales, the measurements of which combined are the measurement of the construct.

Let’s take a look at a research study in which mindfulness was the target construct. Mindfulness is a multifaceted construct and is defined by the following statement as proposed by Baer et al. (2008): “Mindfulness is usually defined to include bringing one’s complete attention to the experiences occurring in the present moment, in a nonjudgmental or accepting way” (Brown & Ryan, 2003; Kabat-Zinn, 2003; Marlatt & Kristeller, 1999).

The label, definition, and a sample item for each of the five factors (or subconstructs) of mindfulness are presented in Table 1.4.

In this example, the mindfulness construct can be measured with the five-factor scale developed by Baer et al. (2008) or other scales that are available in the literature. The researcher can also develop a new scale that is customized to the research context and participants. In the latter case, the researcher has to conduct a scale development project to test the veracity of the conceptual framework first before conducting research using the scale to measure mindfulness.

Knowing that a construct should generally be measured using a scale, it should be noted that a construct *can* also be measured by a single item. For example, if the focus of the survey is on employment status of a population, say adults in Toronto, a question: “Are you currently employed?” will suffice. However, if we want to measure how stressed adult Torontonians are because of the economy, we have two options:

Option #1: We can simply ask a general overall question, “How stressed are you?” and we provide them with a rating scale like a ten-point scale with “1” being *No Stress At All* to “10” being *Very Stressed*.

Option #2: Alternatively, we can ask a number of items pertaining to symptoms (indicators) of stress like “Do you have difficulty relaxing?” “Do you eat quickly?” “Are

Table 1.4 The Five-Factors of Mindfulness (From Baer et al., 2008)

<i>Factor (Subconstruct)</i>	<i>Definition/Description</i>	<i>Sample Item</i>
(1) Observing	Noticing or attending to internal and external experiences, such as sensations, cognitions, emotions, sights, sounds, and smells.	“I notice the smells and aromas of things.”
(2) Describing	Labeling internal experiences with words.	“I am good at finding words to describe my feelings.”
(3) Acting with awareness	Attending to one’s activities of the moment and can be contrasted with behaving mechanically while attention is focused elsewhere (often called automatic pilot).	“I find myself doing things without paying attention.” (R)*
(4) Non-judging of inner experience	Taking a non-evaluative stance toward thoughts and feelings.	“I think some of my emotions are bad or inappropriate and I should not feel them.” (R)*
(5) Non-reactivity to inner experience	The tendency to allow thoughts and feelings to come and go, without getting caught up in or carried away by them.	“I perceive my feelings and emotions experience without having to react to them.”

*NOTE: R = reverse-scored item (higher scores represent higher levels of mindfulness).

you able to concentrate on performing your tasks at work?” and given how the questions are stated, we could provide a two-point *Yes/No* rating scale (also referred to as a dichotomous rating scale or checklist). We then administer this group of questions and compute the average of the item responses (with or without weighting the items based on salience of the items) which we use as a measure of stress. It is because we add up or average the scores across items of a scale that we refer to scales as additive or summated scales. (Item writing and rating scale are discussed in Chapters 8 and 9.)

Other examples of measuring abstract constructs with scales and single items are shown in Table 1.5: The issue of using a single item to measure a construct is further discussed in Chapter 4.

Using a single item to measure a construct is not recommended because constructs are often multifaceted and complex and a single item will not be sufficient to measure it reliably and accurately. (Reliability and validity are discussed in Chapters 12 and 13.) The reason for highlighting this point is that single items to measure abstract attributes have been used in the real world and it is a practice that quantitative researchers should avoid if possible. An exception of sorts may occur when researchers use multiple scales to assess the same construct *and* include a single global item—their purpose may be to assess the *validity of that single item*.

Table 1.5 Single Item and Scale

Topic	Sample Items in Scale	Single Item
Graduate students' satisfaction with their program of study	Satisfaction with - course options - number of courses required - course schedule - quality of instructors - resources...	How satisfied are you with your program of study?
Attitude toward qualitative and quantitative research methods	- One can put aside his or her viewpoints when doing research. - In designing a research, researchers must clearly state what results they expect to obtain.	How strongly do you prefer to use either qualitative methods or quantitative methods for your research?
Linguistic intelligence	- Do you enjoy writing? - Do you love to read?	How intelligent are you linguistically?

1.3.2 Index

A type of measure that is confused with scale is index. Here, we explain the distinction between the two. Scales can be called *reflective measures*, while indices are *formative measures*. The majority of measures in business and the social sciences assume that items *reflect* the level of an underlying latent construct, and so changes in the construct would result in changes in the item response since the construct is seen as *causing* responses to items. For example, items in a depression scale measure the level of depression (depression causes responses to items). But some constructs may be better measured using *observed variables* rather than self-report responses to items. This type of measure is known as *an index or formative measure*; the construct is caused or formed by the variables. Figure 1.2 displays the direction of causal effect for a scale and for an index.

Examples of indices include socio-economic status (SES) (caused by the observed variables of education, income, and occupational prestige), frustration in an experimental

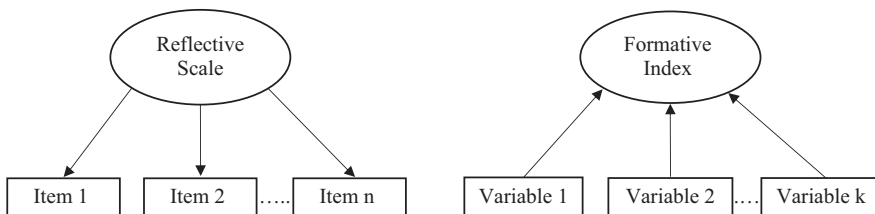


Figure 1.2 Reflective Scale and Formative Index.

setting (caused by withholding food and sleep), or electronic service quality (caused by security/privacy, fulfillment/reliability, website design, customer service, informativeness, and customization) (Blalock, 1964, as cited in Bollen & Davis, 1994/2009; Theodosiou et al., 2019). In Figure 1.3, we show an example of a scale model for a hypothetical family relationship construct and an example of an index model for social status.

Edwards (2010), comparing reflective (scale) and formative (index) approaches, concluded that formative measurement is not viable and recommends that researchers use reflective measurement that achieves the same objectives. Both Edwards (2010) and Howell et al. (2007) suggested designing measures using reflective indicators but note that, for example, requesting self-report of a person's SES tells us their *self-perceived* SES level but may not accurately reflect their *objective* SES level. SES and some other constructs might be better modeled formatively with indicators like occupational prestige, income, level of education, and place of residence.

According to Jarvis et al. (2003), whether a particular set of items are reflective or formative can be judged by (1) the direction of causality between the construct and the variables, (2) the interchangeability of variables, (3) the expected relationship across variables, and (4) the pattern of antecedents and consequences of variables. In a formative model or index, the variables create the construct, so changes in variables are expected to lead to changes in the construct, items are not usually nor are they expected to correlate, and the antecedents and consequences of variables differ (e.g., the antecedents and consequents of “education” and “income” differ). In a reflective model or scale, the construct determines the responses to items, with items that are conceptually

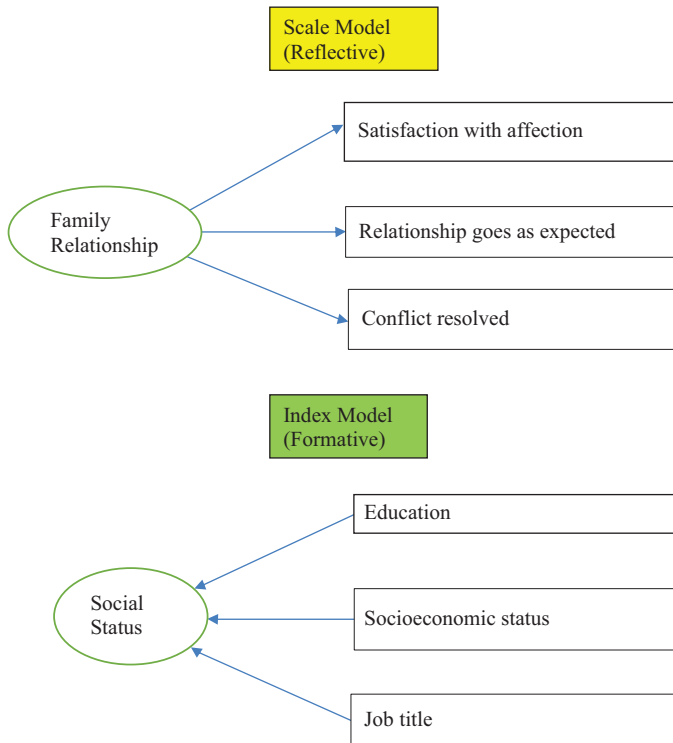


Figure 1.3 Scale and Index Models.

interchangeable. Replacing one item with a similar one does not change the meaning of the construct—say, depression—and is reflected by any item from some universe of items. As long as the items in the scale are equal members and are drawn from the universe of items that reflect the defined construct (depression), what specific items from the universe of items end up in the scale matters little. And, items are expected to correlate and to have the same genesis and consequences (See Chapter 6.).

Does it matter if we view our measure as formative or reflective? Yes, it does. As mentioned, if reflective, we might decide to delete an item that does not correlate with the total, but, if formative, item deletion changes the definition of the construct. And, structural models (as analyzed by structural equation modeling to be discussed in Chapter 13) are specified differently for reflective compared to formative models (van Rooij et al., 2017). This book takes its focus on reflective measures or scales but the process of constructing the survey remains mostly the same whether the measure is a scale or index.

1.4 SURVEY RESEARCH

A survey can be used in experimental research (or program evaluation that is designed to examine the impact of an intervention) to identify problems and needs in needs assessment, and to measure treatment implementation, outcome, and impact. In fact, survey is the most popular data collection tool in training evaluation (with training being the treatment) to measure participants' reaction to the training and/or perception of the effects of training on them. However, experimental research studies that use survey to collect data are not considered as survey research studies *per se*. That is, we differentiate between a survey used as a data collection method and survey research.

As discussed earlier, non-experimental research does not involve any deliberate intervention (e.g., we might survey current teachers' student-centered instruction practices). Not all non-experimental research studies are survey research studies. First of all, non-experimental studies that use only qualitative methods with a very small sample size are not survey research. Non-experimental studies that use quantitative data collection methods other than surveys (e.g., observation, examination of records and tests) are technically not survey research either. However, a non-experimental study (research that does not involve any intervention or treatment that is manipulated by researchers) that is conducted to find out (through self-report) people's opinion about the current economic crises, or teachers' change in attitude toward student-centered instruction and the education reform movement, is survey research. (See Table 1.3 for more example.)

Simply stated, *survey research (also referred to as descriptive research) is a type of non-experimental research in which administering surveys is the primary or sole data collection procedure*. Consequently, development of survey plays such a significant role in survey research that the distinction between survey research and survey development is blurry. Survey research can be used to assess population characteristics (e.g., political party affiliation or mindfulness), to understand how a certain group of people feel about certain issue (e.g., abortion or climate control), and/or to determine the associations between variables of interest (e.g., education level and mindfulness). A distinction should be made between survey research and research using surveys.

In methodological research, survey can be the focus of experimental manipulation for the purpose of improving survey usage. For example, we can create two versions of a scale, in one we group the items by theme and in the other we randomly present the items. We then randomly assign each of two versions to half of the sample, and see what the effects of

grouping items are. Although survey is the focus of the study, this type of research is experimental and hence is not considered as survey research even though we would use surveys to collect our data. This type of research may be referred to as research on survey or scale. (Research involving scales is discussed below.)

The basic procedure for conducting survey research involves multiple stages: (1) determining the purpose of the survey in regard to intended interpretation and use of results, (2) developing a blueprint of the survey, (3) writing or finding appropriate items and reviewing and revising the items and the questionnaire, (4) pretesting the first completed draft of the questionnaire and further revising instruction and items if necessary, (5) administering the final version of the questionnaire to collect data, (6) analyzing and interpreting data, and (7) drawing conclusions and reporting findings. We will further describe the survey development and validation process later on.

1.4.1 Survey Research Design

Survey research uses different designs to address different research questions. These designs differ by the *frequency* of survey administration, and the *composition of respondents*. More specifically, these designs can be differentiated along the following three dimensions stated as questions:

In the same study, whether or not

- (1) The survey is administered across time? (Across Time)
- (2) The same population or different populations will be surveyed? (Population)
- (3) The same or a different sample of a population will be surveyed? (Sample)

On the basis of these three dimensions, survey research designs can be either a *single-wave* or a *multiple-wave* design; multiple-wave designs can either be cross-sectional or longitudinal designs; longitudinal designs can be further divided into trend, cohort, and panel designs. All the multiple-wave designs aim to measure *changes* in a variable or construct (e.g., changes in teaching effectiveness or empathy). These various designs and their defining characteristics pertaining to these three dimensions can be seen in Table 1.6.

Table 1.6 Survey Research Design

<i>Survey Research Design</i>	<i>Across Time Yes or No?</i>	<i>Population Same or Different?</i>	<i>Sample Same or Different?</i>
(1) Single wave (or point)	No	Same	Same
(2) Multiple waves (or points)	–	–	–
(a) Cross-Sectional (or pseudo panel)	No	Different	Different
(b) Longitudinal	–	–	–
1. Trend	Yes	Different	Different
2. Cohort	Yes	Same	Different
3. Panel	Yes	Same	Same

Now we can combine the survey research designs with the research foci to create the following Table 1.7:

Table 1.7 Survey Research Design and Research Focus

Survey Research Design	Research Focus		
	Status	Group Difference	Relationship
(1) Single wave (or point)	–	–	–
(2) Multiple waves (or points)	–	–	–
(a) Cross-Sectional (or pseudo panel)	–	–	–
(b) Longitudinal	–	–	–
1. Trend	–	–	–
2. Cohort	–	–	–
3. Panel	–	–	–

These three research foci can be studied with the single-wave and the four multiple-wave designs. Note again that all the multiple-wave designs examine *changes* in the three research foci. Here are sample research questions of the three foci for single-wave design and for all the four multiple-wave designs:

Single-wave design:

- How emotionally intelligent are secondary teachers? (Status)
- To what extent do male and female secondary teachers differ in EI? (Group difference)
- What is the relationship between secondary teachers' EI and teaching performance? (Relationship)

Multiple-wave designs:

- To what extent does the EI of secondary teachers *change* across time? (Status)
- To what extent do sex differences in EI of secondary teachers *change* across time? (Group difference)
- To what extent does the relationship between secondary teachers' EI and teaching performance *change* across time? (Relationship)

To carry out these research studies, the teachers will be recruited in the following way for the different designs:

- A sample of secondary teachers (single wave);
- A sample of newly hired secondary teachers and a sample of experienced secondary teachers recruited at the same time with data collected from experienced teachers representing the anticipated future performance of newly hired teachers (Cross-sectional);

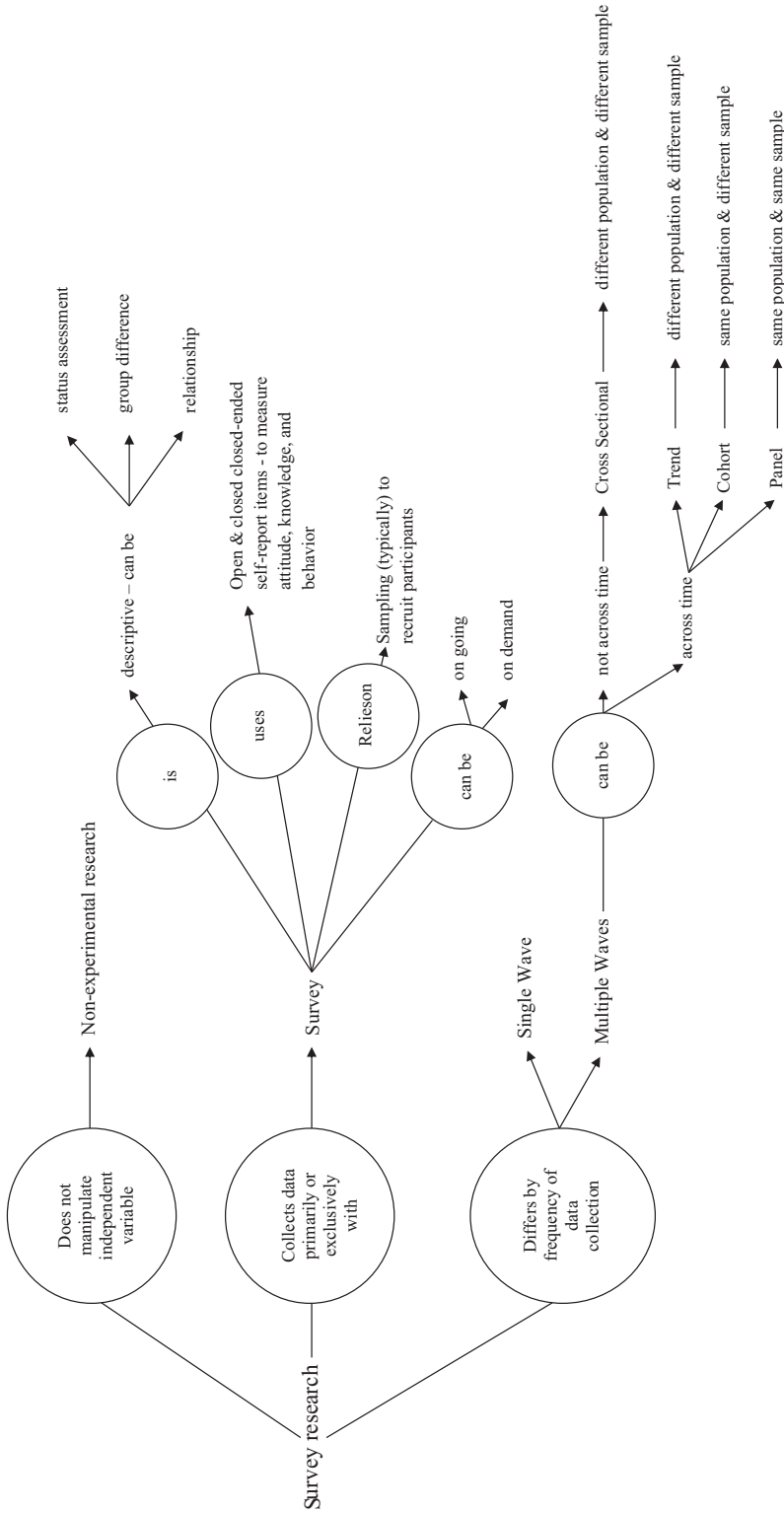


Figure 1.4 A Concept Map of Survey Research.

- A sample of secondary teachers graduated in 2020 and a different sample of secondary teachers graduated in 2022 (Trend);
- A sample of secondary teachers graduated in 2020 and surveyed in 2020, and a different sample of secondary teachers also graduated in 2020 but surveyed in 2022 (Cohort);
- A sample of newly hired secondary teachers in their first, third, and fifth years of teaching (Panel).

To summarize, in Figure 1.4, the defining characteristics of survey research and the various survey research designs are depicted in a concept map.

1.4.2 Survey Research with Scales

Survey research studies that contain one or more constructs (as measured by scales) can be grouped into the following three categories:

1. *Scale development research.* The purpose of this type of research is to construct and validate a new scale, for example, developing a teacher efficacy scale.
2. *Research-on-construct.* Research of this type focuses on studying a “primary” construct (e.g., mindfulness) using either experimental or non-experimental designs in regard to status (e.g., how empathetic are education professors?), relationship (e.g., what is the relationship between empathy and narcissism?), or difference (e.g., what is the difference in instructional practices between secondary teachers with high degree of empathy and secondary teachers with low degree of empathy?). This type of research study is designed to examine an established construct and can be considered as further assessing the validity of a construct beyond that reported in seminal scale development research (see below and more on that in later chapters).
3. *Research-with-construct.* In this type of research, the construct is included but it is a “secondary” variable in the research. For example, using empathy as an example, a research study that aims to study the effects of a training program on student-centered instruction on students’ “motivation to learn,” which is the “primary” construct, may include empathy as a *mediating* variable that causally relates training to motivation to learn. In another study that examines physicians’ ability to self-assess their clinical interview skills (which is the primary variable) and clients’ satisfaction with their performance, empathy may be included as a secondary variable, as a *moderating* variable that affects the relationship between physicians’ ability and clients’ satisfaction.

We can say with some confidence that antecedent and consequent variables are usually the primary constructs of interest and mediating/moderating variables are not. Although a mediating (or moderating) variable usually acts as a secondary variable, sometimes it can be the primary variable. For example, researchers may want to study the effect of frustration (as the primary variable) on decision making using an experimental design. They may induce frustration by putting participants in a confined room alone for hours before asking them to make decisions.

Wait time (proxy variable) → frustration (primary variable) → decision making