1 Why we Conduct Research?

2 The Research Process

3 Phase A: Deciding What to Research

4 Phase B: Planning a Research Study

Introduction to Quantitative Research Methods in Educational Research

S

N

A

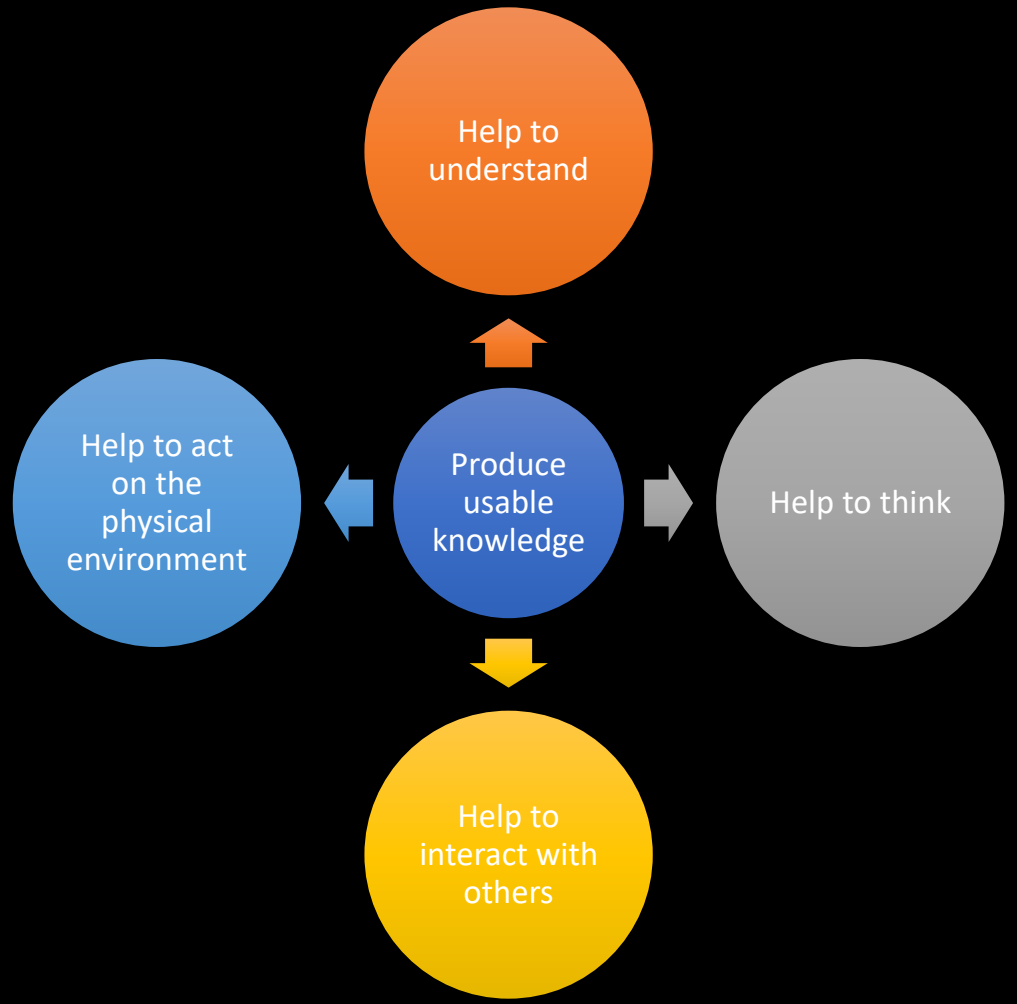
Siti Noormi Alias
Senior Lecturer
Department of Professional Development and Continuing Education
Faculty of Educational Studies
Universiti Putra Malaysia

Learning Outcomes

At the end of the session, participants will be able to:

1. Understand the details about the steps of quantitative research methodology; its types, strengths and weaknesses;
2. Generate research ideas and formulate relevant research questions; and objectives
3. Identify the key points regarding different types of quantitative research.

Why We Conduct Research?



3 Basic types of Epistemological Belief

Empiricism/ Sensationalism

Sources of and justifications for knowledge claims come from observation through the senses

Direct experience provides the data or evidence for developing and then answering our questions.

Without direct observation, research questions
simply would not occur to us

Rationalism/ Idealism

Knowledge arises from and is verified by reasoning using internal mental categories

Manifests as a preference for deductive learning.

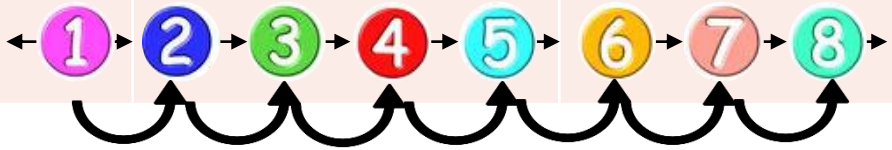
We begin with a truth or premise, then, through logical deduction, generate new truths or elaborate on that premise; this results in a testable hypothesis.

Sociocultural relativism

Knowledge is shaped by the specific social and cultural circumstances of those making knowledge claims

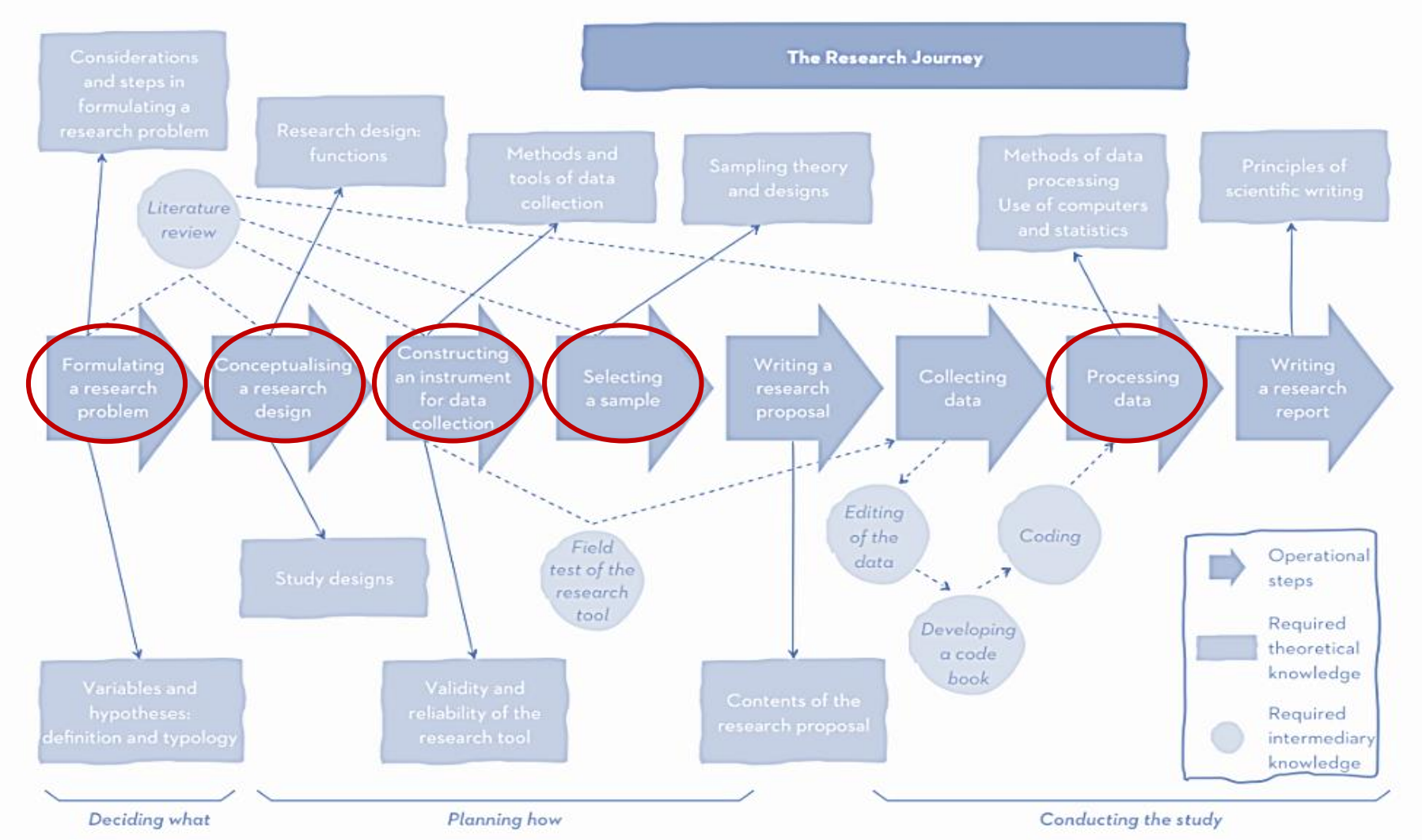
How do our socially and culturally mediated preferences shape what we know, what and how
we investigate, and how we go about turning the results of our investigations into usable knowledge in context?

The Research Process

Phase	PHASE A	PHASE B	PHASE C
Main task	DECIDING ↓ WHAT? (research questions to answer?)	PLANNING ↓ HOW? (to gather evidence to answer the research questions?)	CONDUCTING ↓ COLLECTING (the required information)
Operational steps/ research journey			

Source: Kumar (2019, p. 43)

Quantitative Research Process



Source: Kumar (2019, p. 43)

Phase A: Deciding what to research

Step 1: Formulating a Research Problem

- Identify what you intend to research
- Decide what you want to find out about
- Be specific
- Things to consider
  - Financial resources
  - Time available
  - Supervisor’s knowledge & expertise

Types of Research Questions

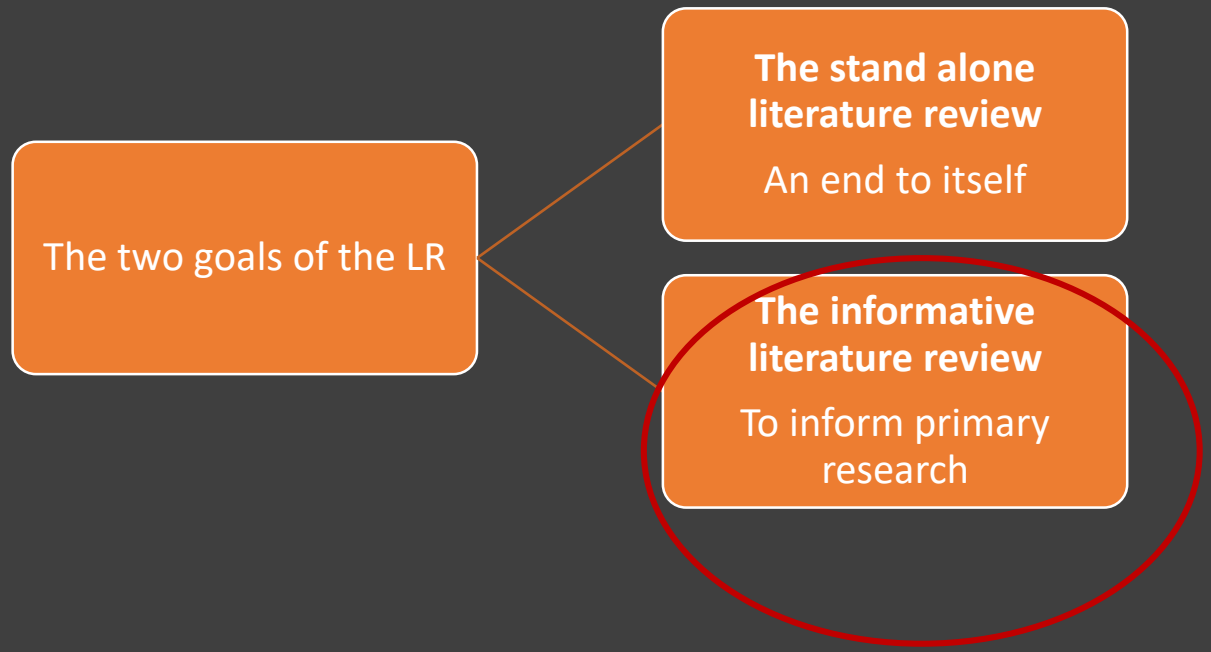
1. ***Descriptive*:** When a study is designed primarily to describe what is going on or what exists. Public opinion polls that seek only to describe the proportion of people who hold various opinions are primarily descriptive in nature. For instance, if we want to know what percent of the population would vote for an UMNO or a Perikatan Nasional in the next general election, we are simply interested in describing something.
2. ***Relational*:** When a study is designed to look at the relationships between two or more variables. A study that compares what proportion of males and females say about work-to-family or family-to-work conflict while working from home during Covid-19 Pandemic is essentially studying the relationship between gender and work-family conflict tendencies.
3. ***Causal*:** When a study is designed to determine whether one or more variables (e.g., a program or treatment variable) causes or affects one or more outcome variables. If we did a public opinion poll to try to determine whether a recent political advertising campaign changed voter preferences, we would essentially be studying whether the campaign (cause) changed the proportion of voters who would vote UMNO or Perikatan Nasional

   (effect).

Why it Must be a Quantitative Research?

- It involves quantification of “something”
  - research question to which we want to find an answer
  - a hypothesis we want to either support or disprove
- The ‘something’ we wish to quantify can take many different forms and be of different kinds
- Research undertaken in education concerned with aspects related to human behavior
  - provide descriptions of students’ achievement variation between different intervention or treatment.
  - Evaluate the impact of a new pedagogic approach in students’ learning and its assessments
- Quantification involves
  - Descriptions
  - Comparisons
  - Associations
  - Predictions
  - Cause and effect

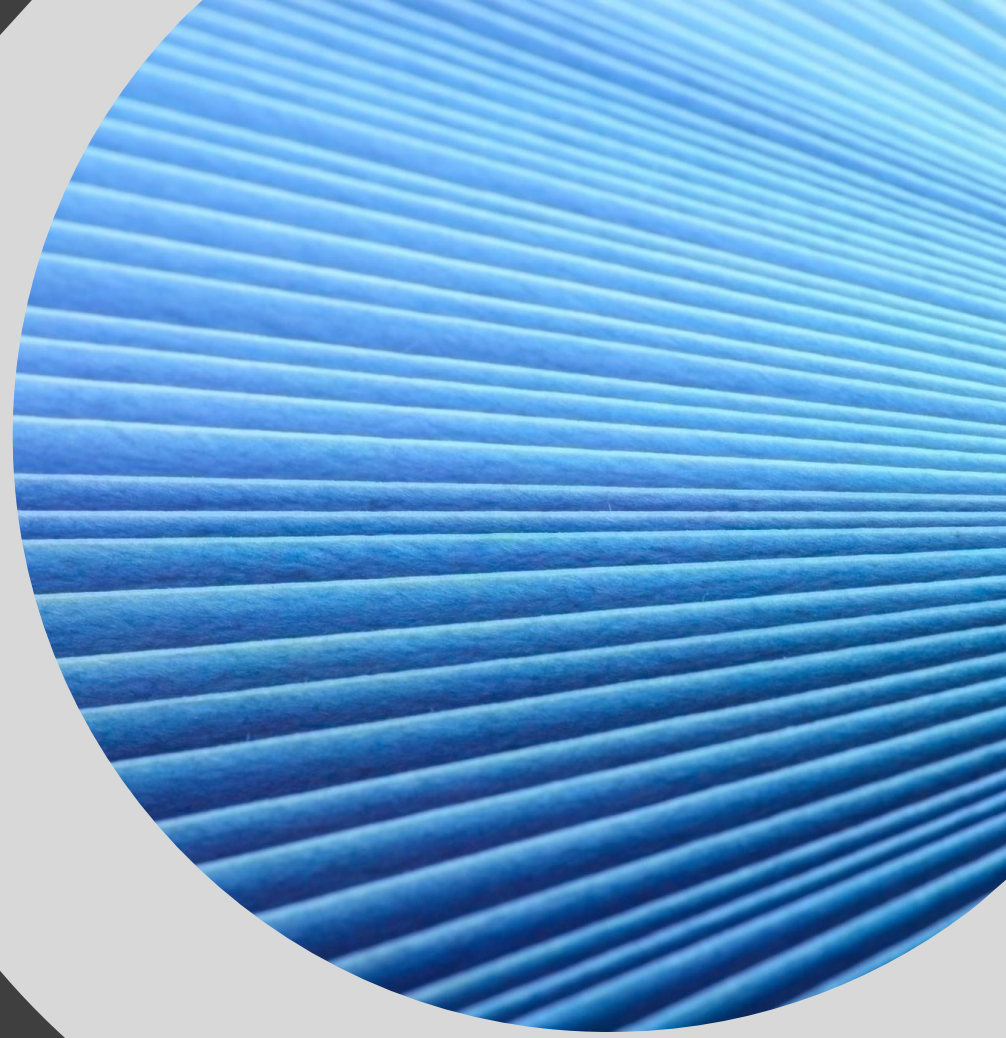
Literature Review



Reasons for Conducting Literature Review

Topic-focused reasons

To Inform Your Topic	To Narrow Your Topic	To Provide a New Lens to Your Topic
Rationalize the significance of a topic	Give focus to a topic	Synthesize and gain a new perspective on a topic
Avoid unintentional and unnecessary replication	Acquire and enhance language associated with a topic	Distinguish exemplary research
Identify key research on a topic, sources, and authors		Make a new contribution on a topic
Identify the structure of a component in a topic		Establish context for author's own interest
Define and limit the research problem		
Identify key landmark studies, sources, and authors		



…cont.

Method-Driven reasons

To Explore New Methods

- Identify philosophical sciences and assumptions used by the authors
- Identify the theoretical, conceptual, and/or practical frameworks used by the authors
- Identify the procedures (e.g., sample size, research design, data collection instruments, and/ or data analysis techniques) used by authors

…cont.

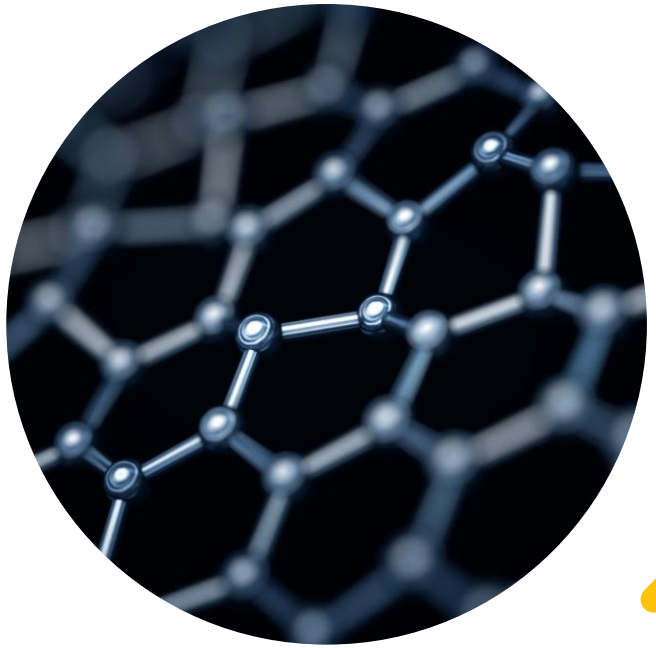
Connection-focused reasons

To make Interconnections with Your Topic

- Identify relationships between theory/concepts and practice
- Identify contradictions and inconsistencies
- Identify relationships between ideas and practice
- Identify strengths and weaknesses of the various research approaches that have been utilized

To Make Outerconnections with Your Topic

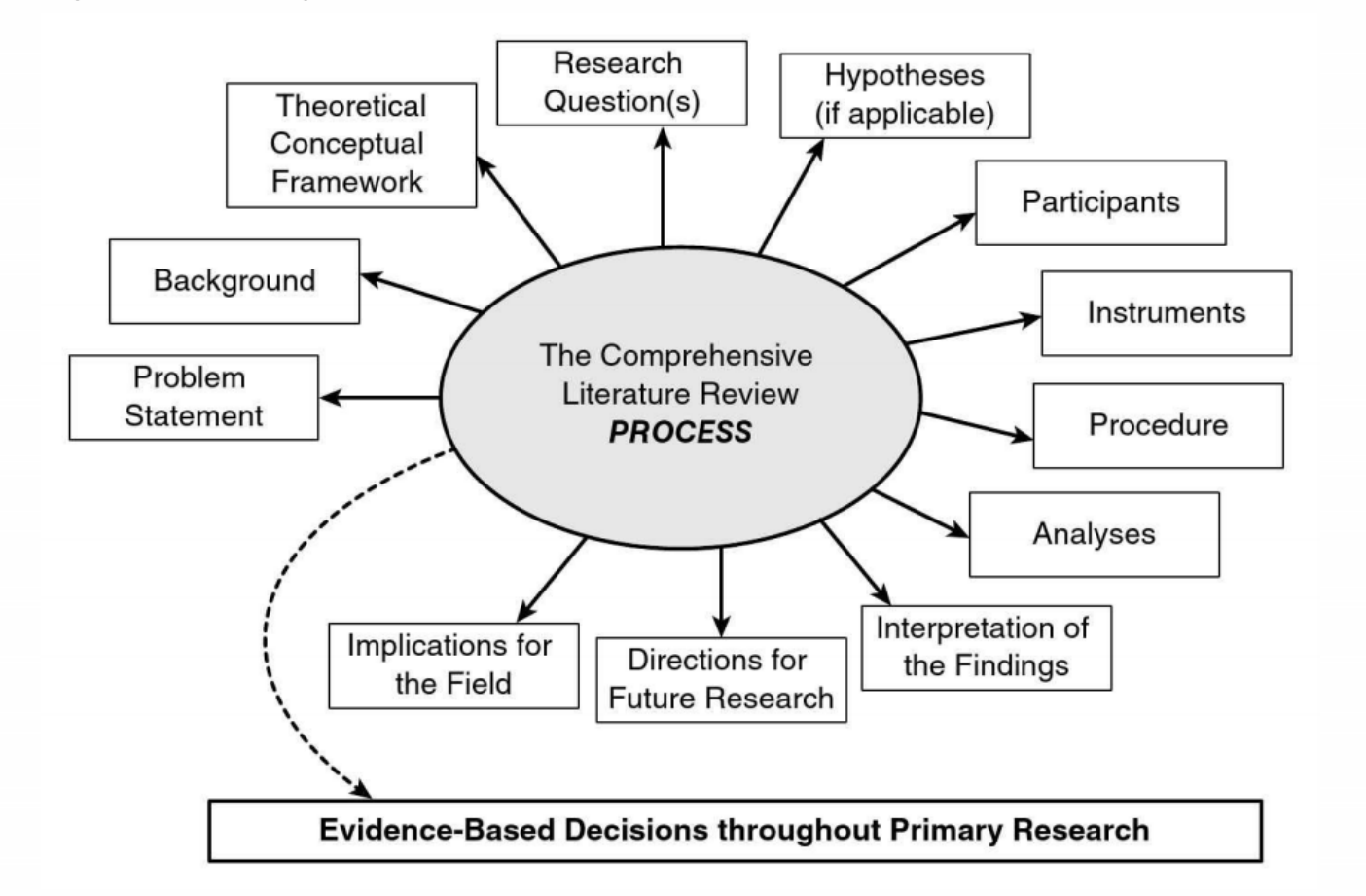
- Distinguish what has been researched and what needs to be researched
- Evaluate the context of a topic or problem
- Bridge the identified gaps on a topic
- Place the research in a historical context
- Provide rationale for research hypotheses
- Form basis for justifying significance of target study
- Identify the scope of the author's investigation
- Provide avenues for future research
- Facilitate interpretation of study results



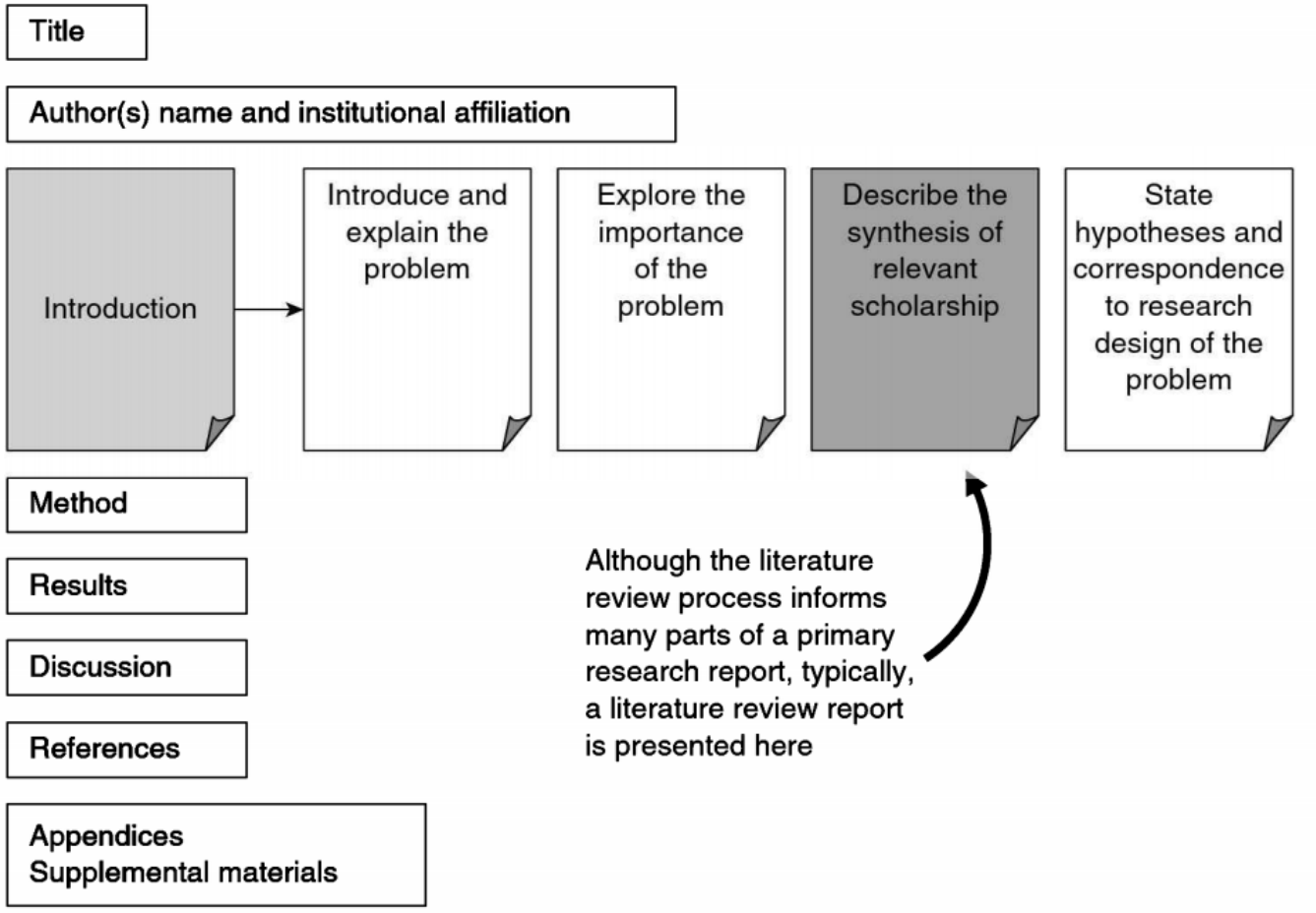
The Literature Review taxonomy of Objective

Literature Review Objective	Cognitive Objective	Affective Process
Highest Level **SYNTHESIZE**	Arranging Comparing Contrasting Translating Categorizing Interpreting Deriving Extrapolating Producing	***Organizing*** Integrating different concepts, theories, findings, and evidences from multiple pieces of information, and accommodating them within his/her own schema; comparing, contrasting, relating, and expanding on what has been learned ***Characterizing*** Holding a particular value or belief that now exerts influence on his/her review, so that it becomes a meta-representation—i.e., a coherent combination of representations stemming from each piece of relevant information
EVALUATE	Assessing Supporting Refuting Justifying	***Valuing*** Attaching a value to each piece of information
ANALYZE	Using Applying	***Responding*** Actively participating in the literature review process; the reviewer not only passively reviews the information but also reacts to it in some way
Lowest Level **SUMMARIZE**	Knowing Understanding Describing	***Receiving*** The lowest level; passively reviewing the information; without this level, no understanding of the literature can occur

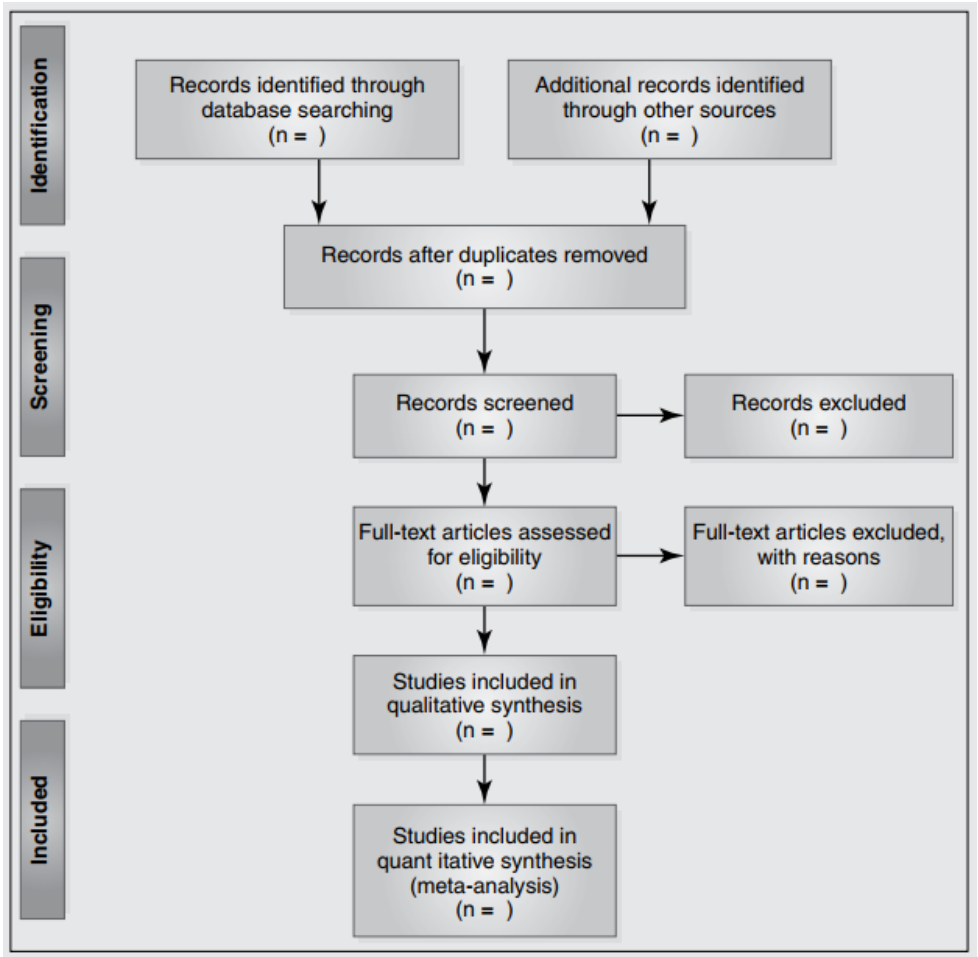
Seven Step Model to Inform Primary Research Areas



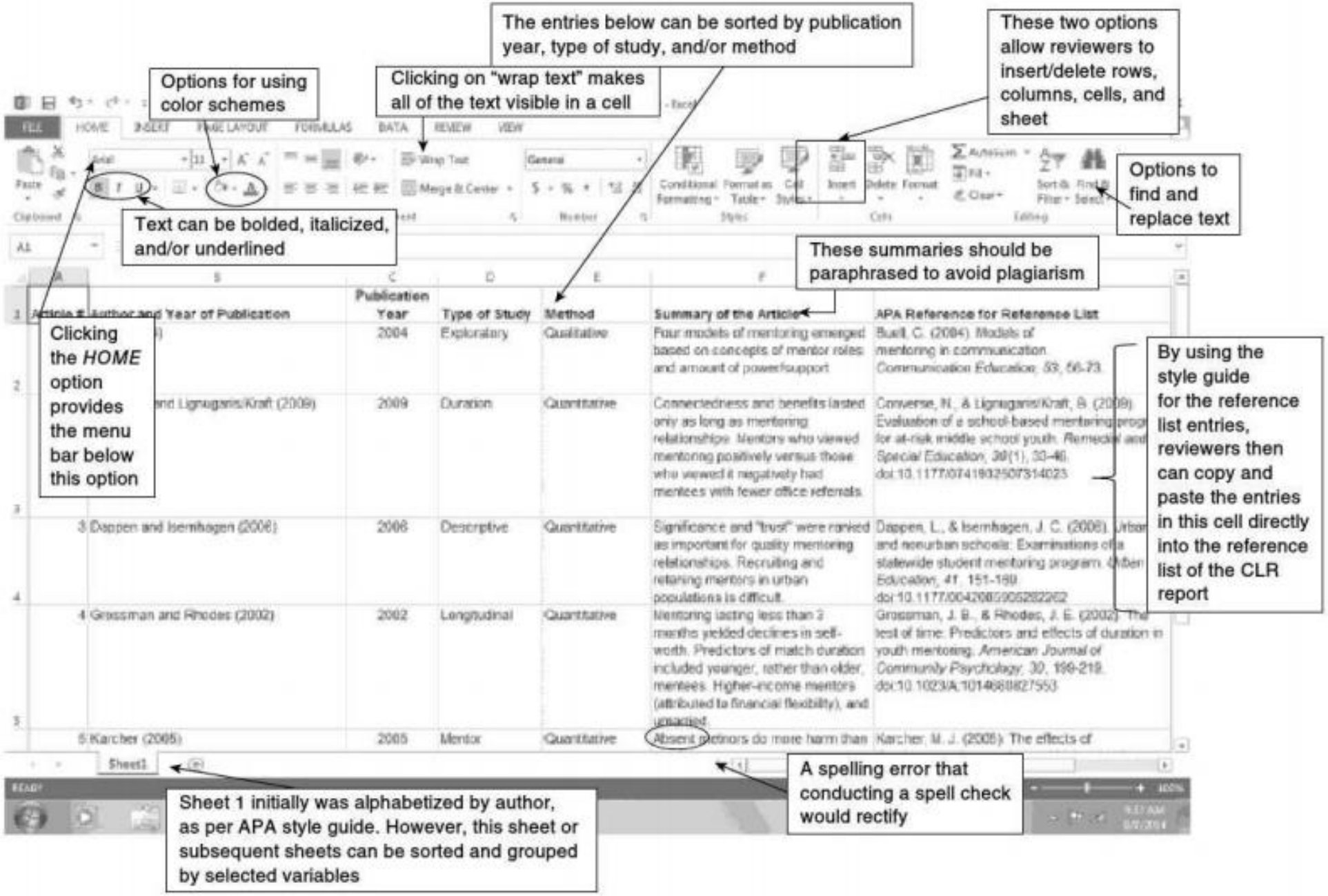
A Map of the Sections of a Typical Primary Research Report

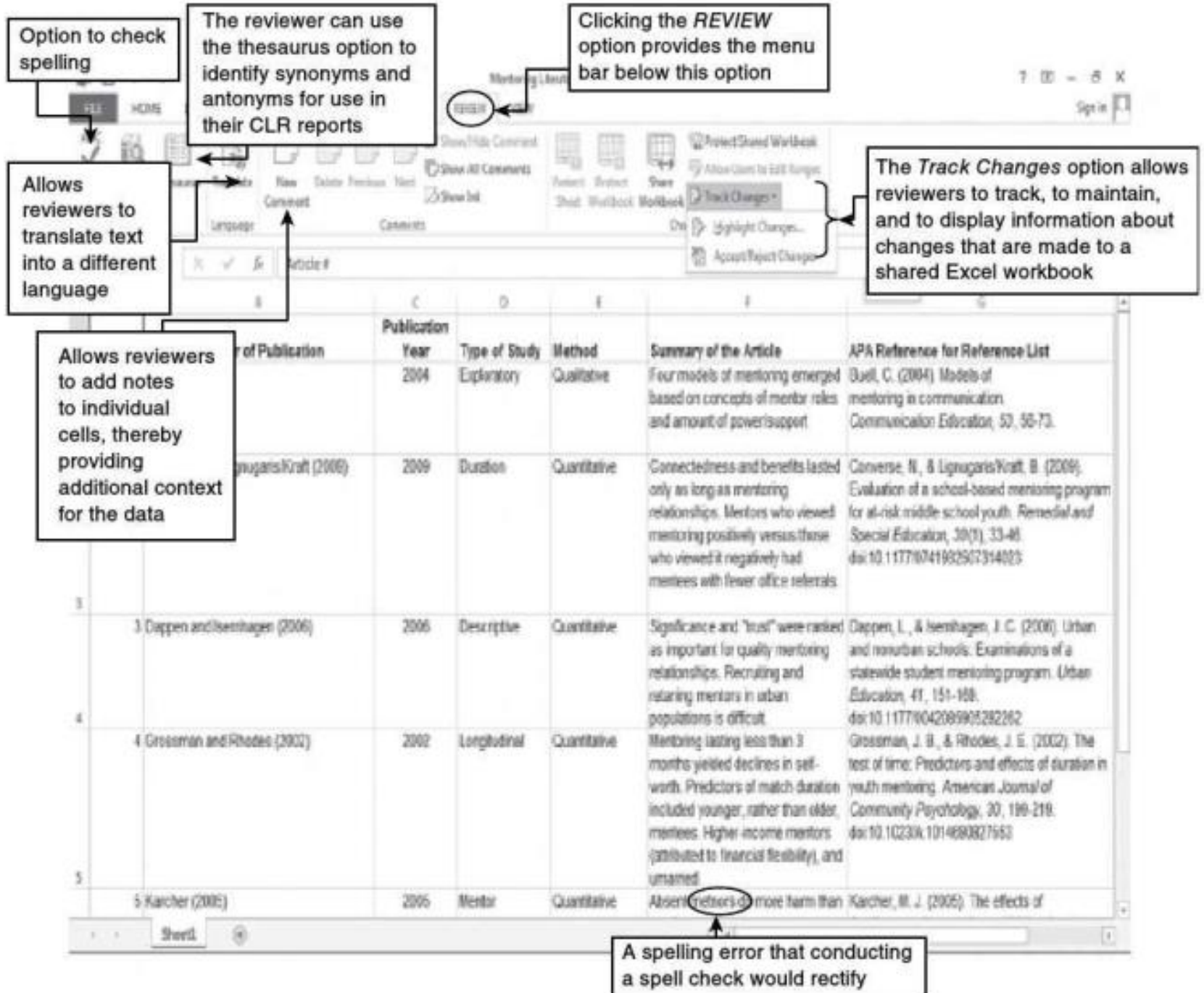


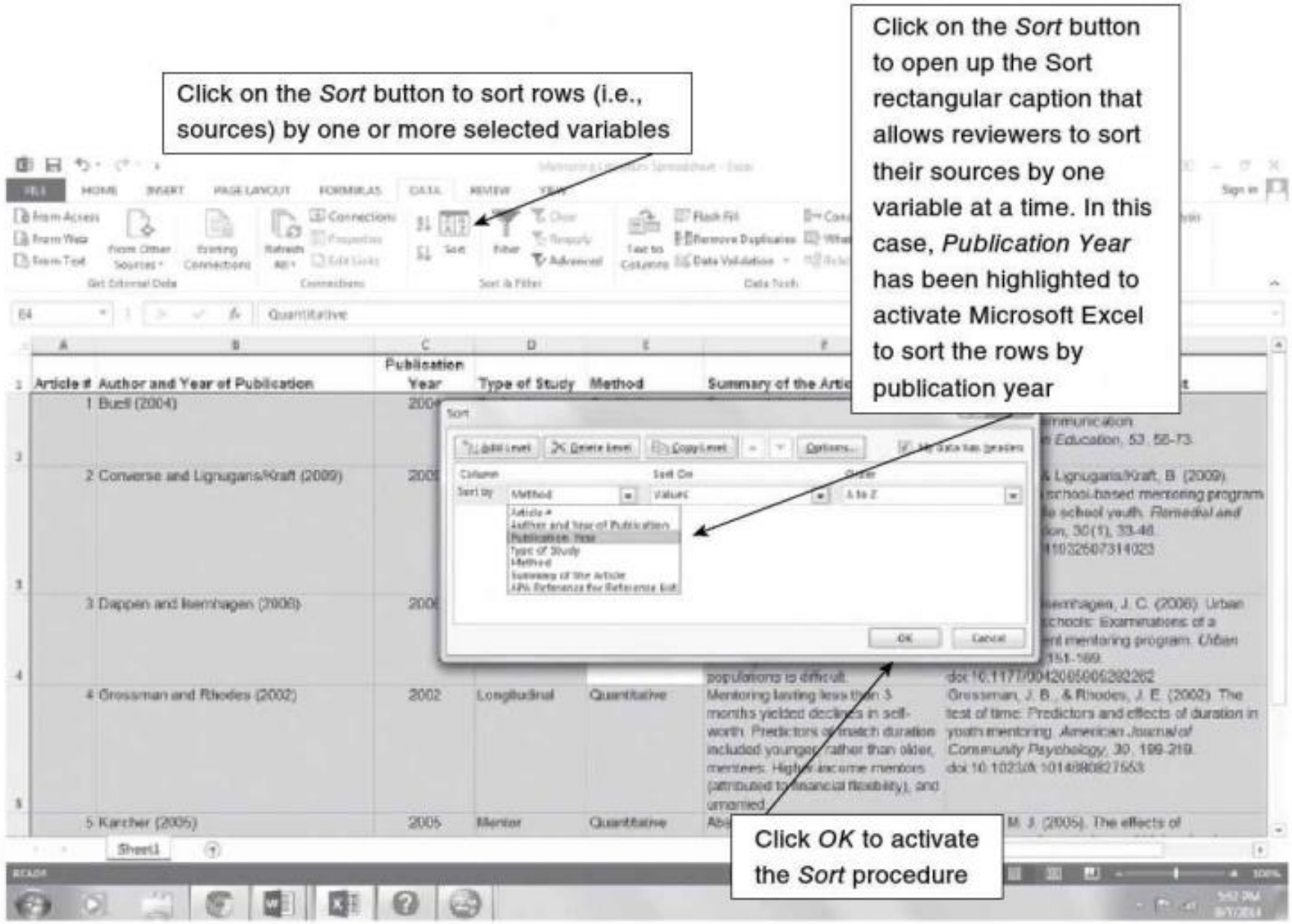
PRISMA Flowchart for Tracking Source Documents throughout the Review Process



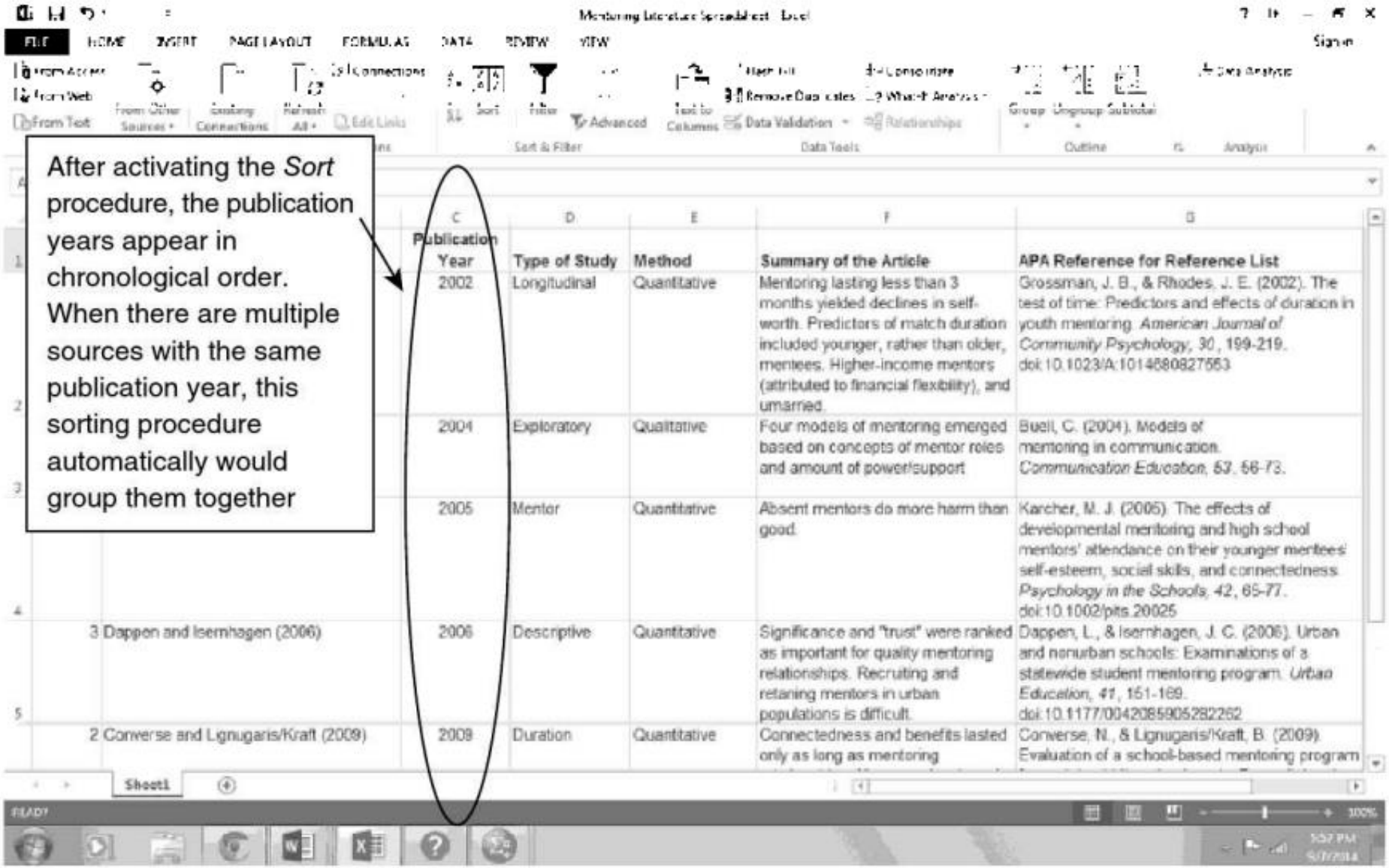
Literature Review Matrix in Excel




Option to check spelling
The reviewer can use the thesaurus option to identify synonyms and antonyms for use in their CLR reports
Clicking the *REVIEW* option provides the menu bar below this option
Allows reviewers to translate text into a different language
The *Track Changes* option allows reviewers to track, to maintain, and to display information about changes that are made to a shared Excel workbook
Allows reviewers to add notes to individual cells, thereby providing additional context for the data
A spelling error that conducting a spell check would rectify

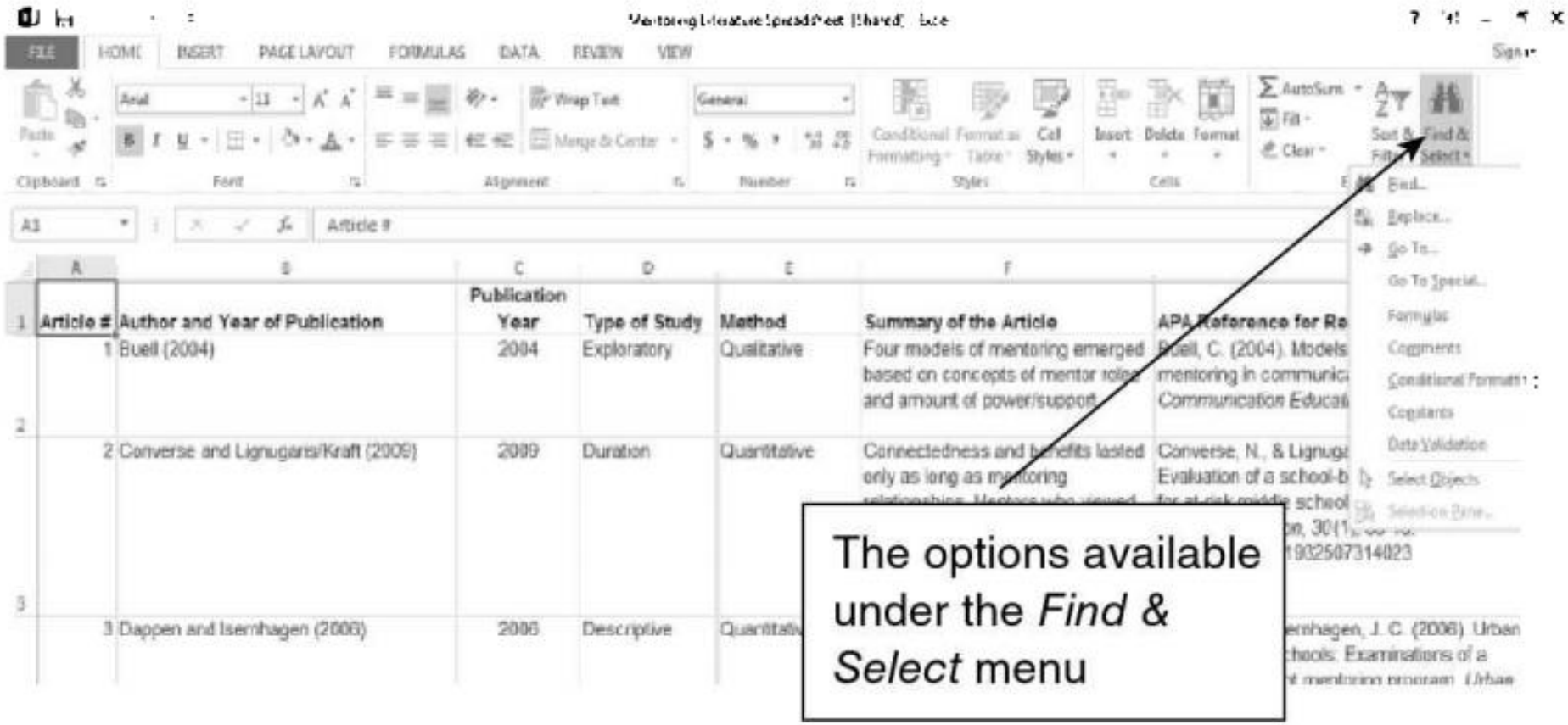


Screenshot showing how to sort sources by a variable

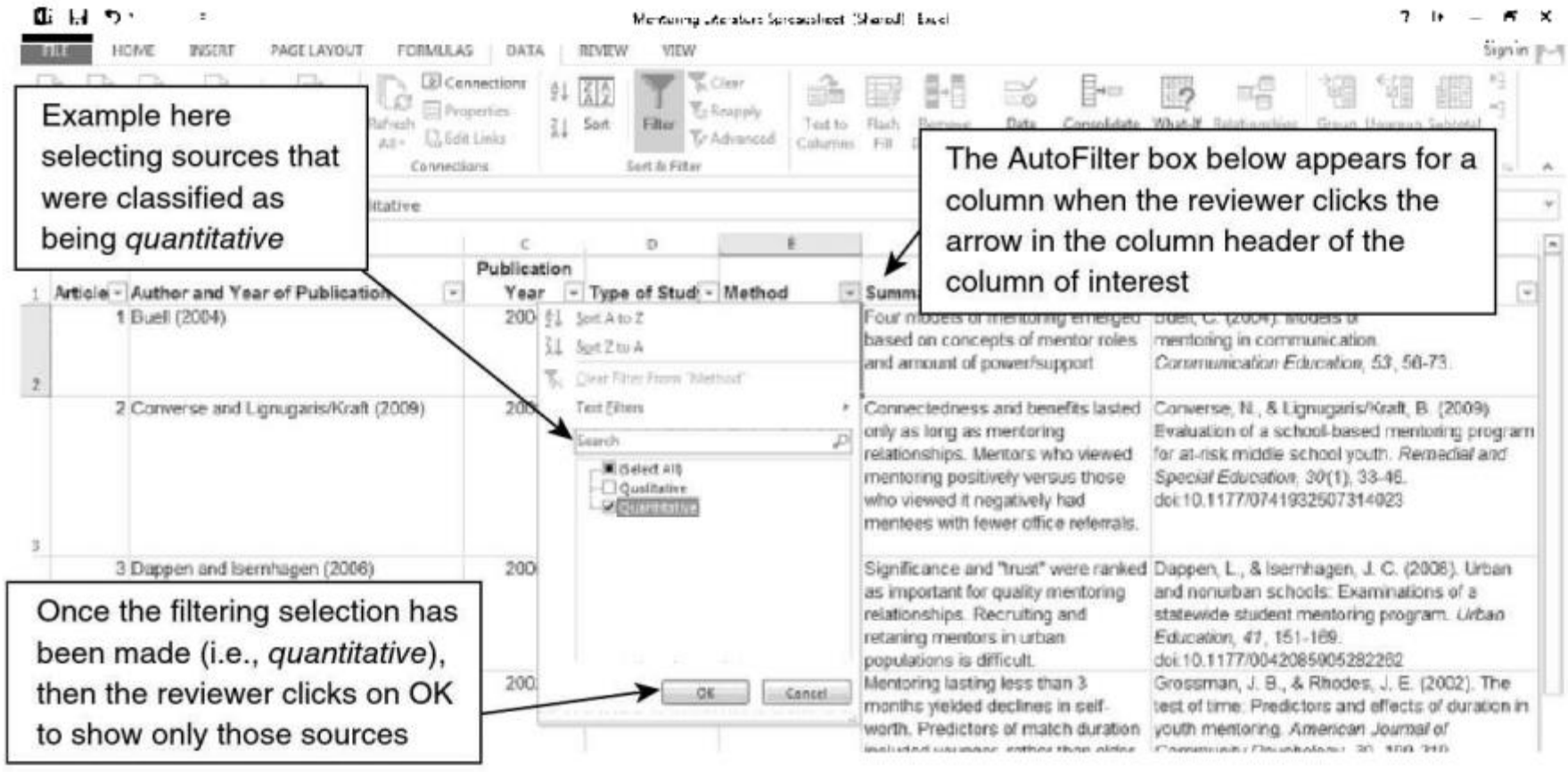


		C	D	E	F	G
		Publication Year	Type of Study	Method	Summary of the Article	APA Reference for Reference List
		2002	Longitudinal	Quantitative	Mentoring lasting less than 3 months yielded declines in self-worth. Predictors of match duration included younger, rather than older, mentees. Higher-income mentors (attributed to financial flexibility), and unmarried.	Grossman, J. B., & Rhodes, J. E. (2002). The test of time: Predictors and effects of duration in youth mentoring. *American Journal of Community Psychology, 30*, 199-219. doi:10.1023/A:1014680827553
		2004	Exploratory	Qualitative	Four models of mentoring emerged based on concepts of mentor roles and amount of power/support	Buell, C. (2004). Models of mentoring in communication. *Communication Education, 53*, 56-73.
		2005	Mentor	Quantitative	Absent mentors do more harm than good.	Karcher, M. J. (2005). The effects of developmental mentoring and high school mentors' attendance on their younger mentees' self-esteem, social skills, and connectedness. *Psychology in the Schools, 42*, 65-77. doi:10.1002/pits.20025
	3 Dappen and Isernhagen (2006)	2006	Descriptive	Quantitative	Significance and "trust" were ranked as important for quality mentoring relationships. Recruiting and retaining mentors in urban populations is difficult.	Dappen, L., & Isernhagen, J. C. (2006). Urban and nonurban schools: Examinations of a statewide student mentoring program. *Urban Education, 41*, 151-169. doi:10.1177/0042085905282252
	2 Converse and Lignugaris/Kraft (2009)	2009	Duration	Quantitative	Connectedness and benefits lasted only as long as mentoring	Converse, N., & Lignugaris/Kraft, B. (2009). Evaluation of a school-based mentoring program

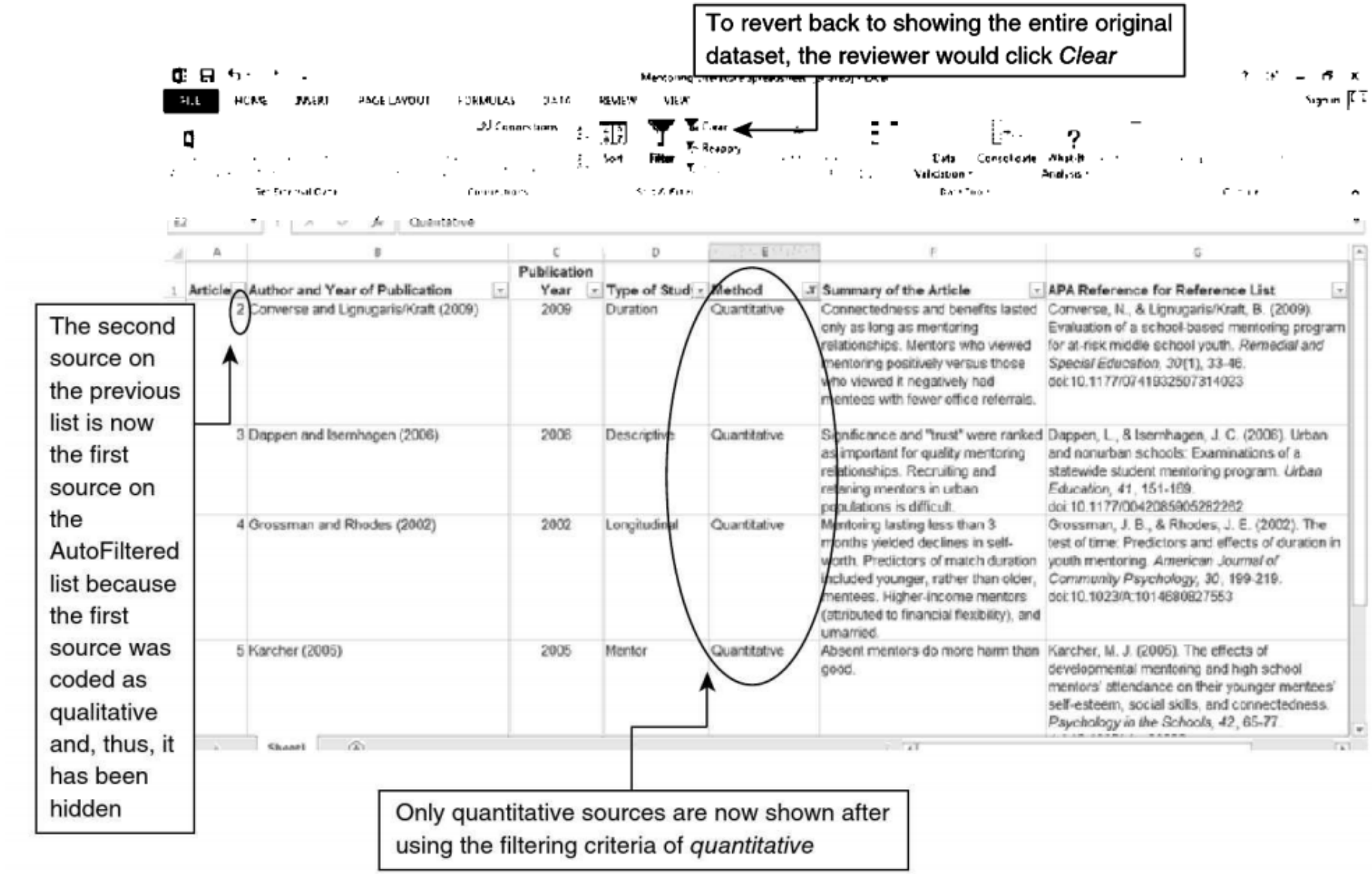
Screenshot showing the sources sorted by publication year (i.e., in chronological order)



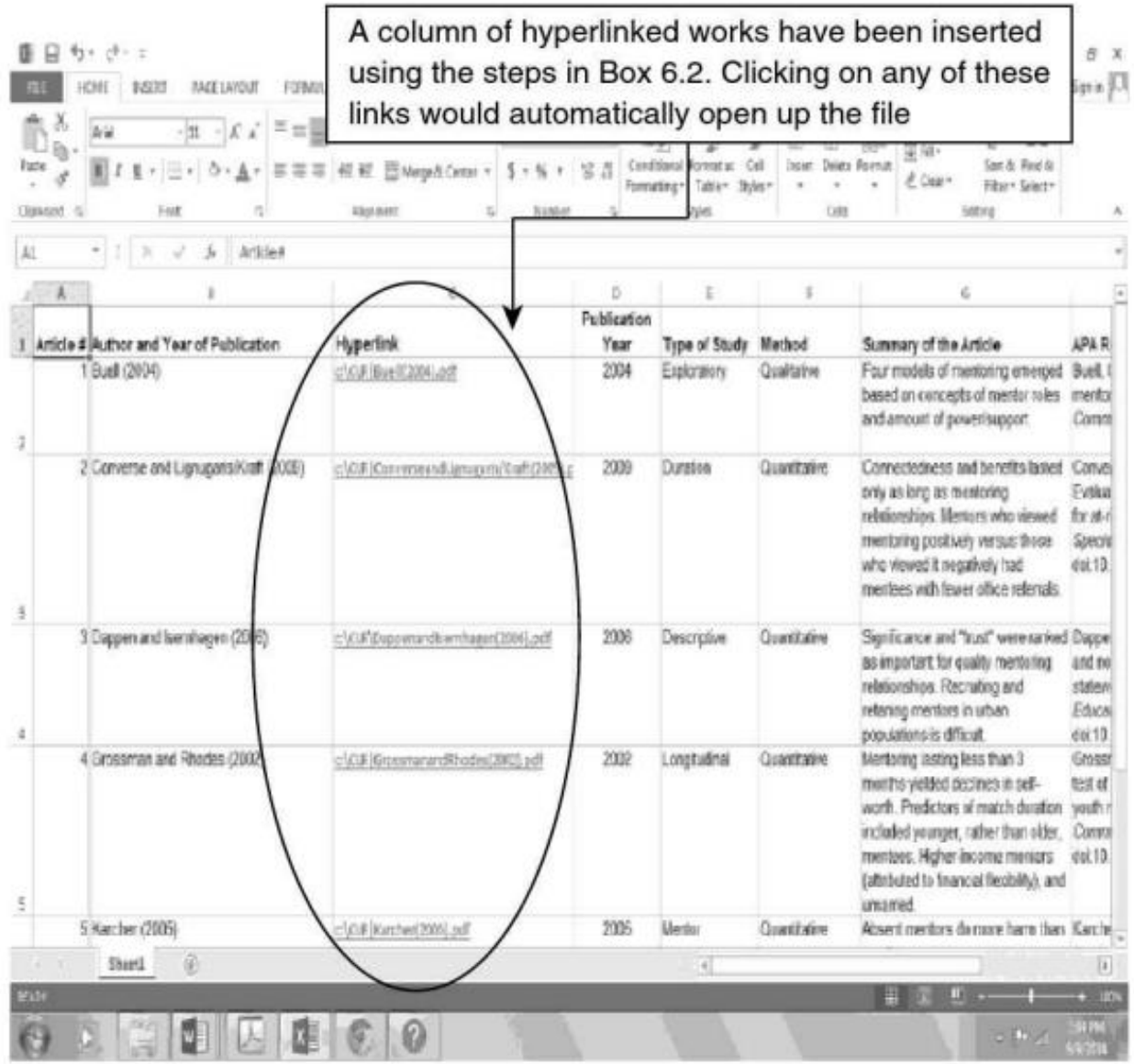
Screenshot showing the *Find & Select* option available to a Microsoft Excel spreadsheet



Screenshot showing the AutoFilter option used to find text in one or more columns of data



Screenshot showing only the quantitative sources after using the filtering criteria



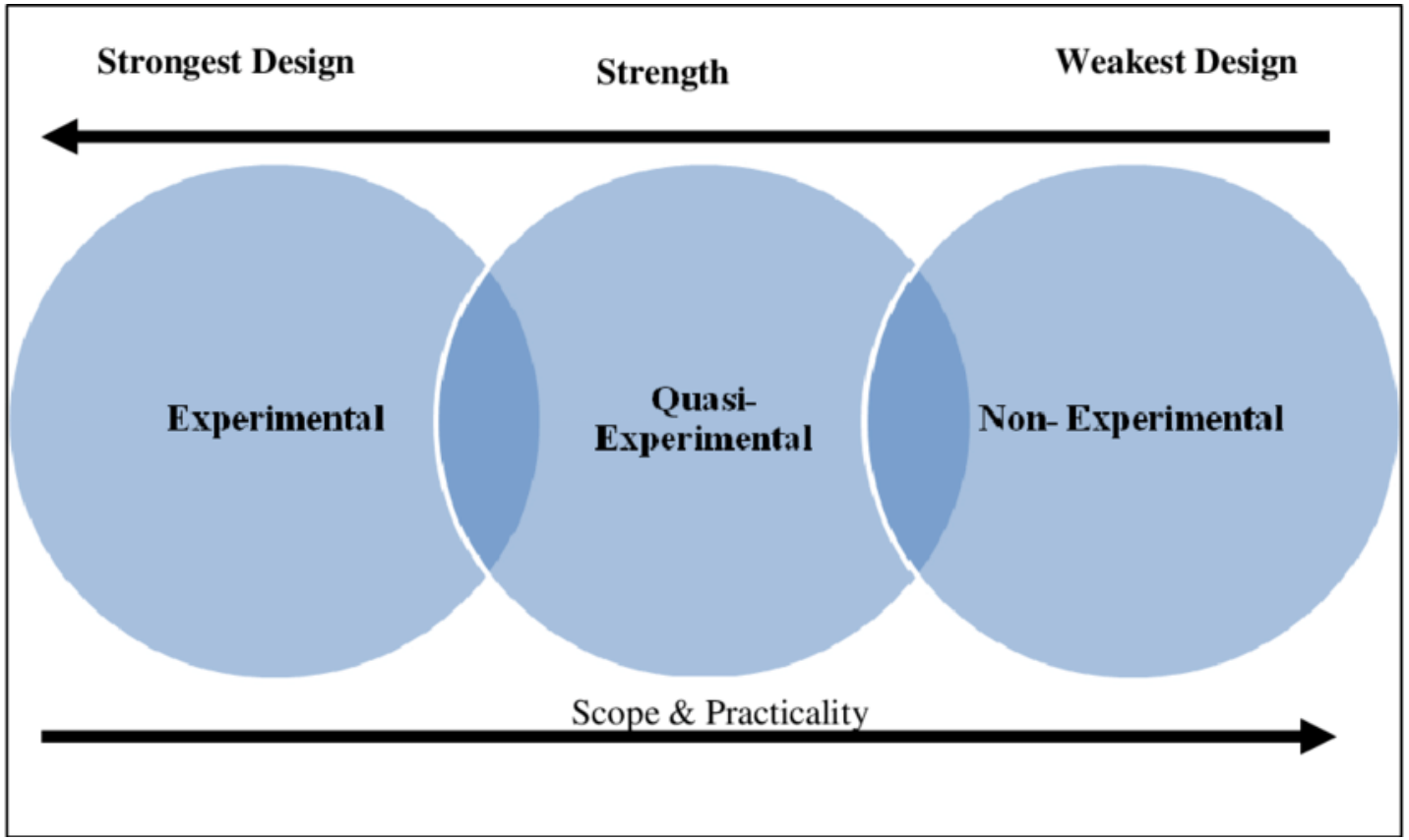
Screenshot showing an example of a column that contains the hyperlinks

Phase B: Planning a Research Study

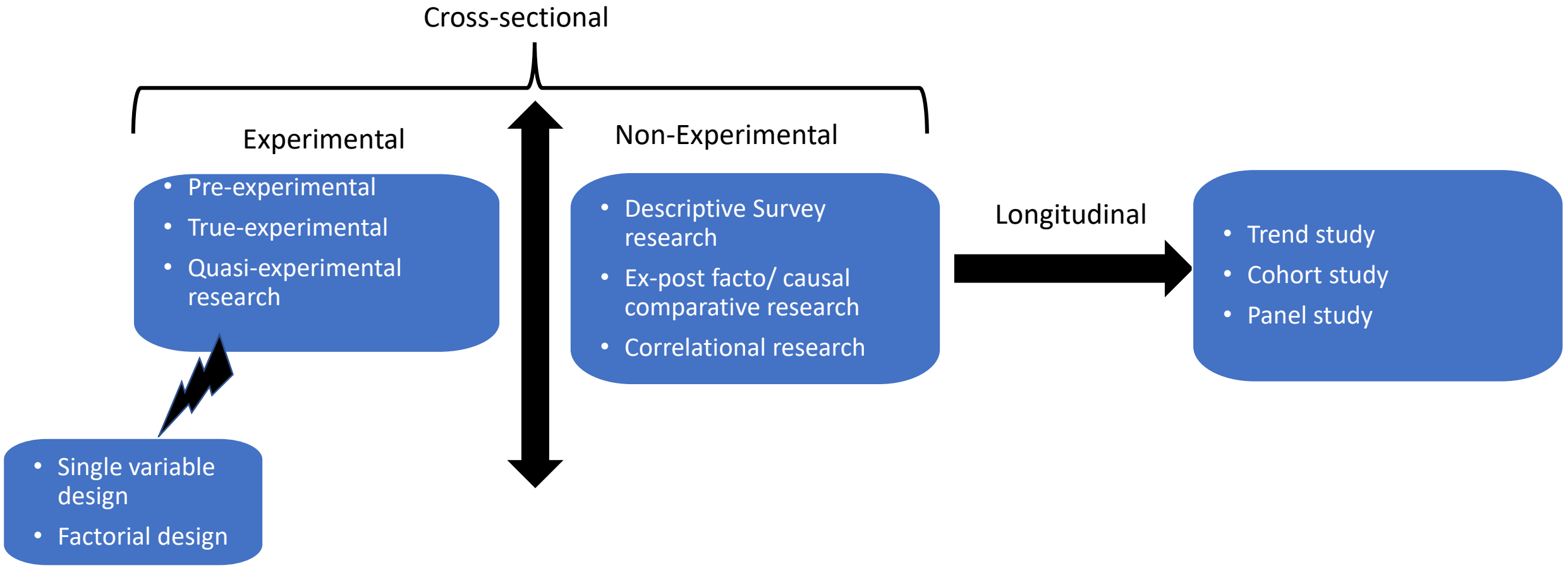
Step 2: Conceptualizing a Research Design

- Select an appropriate research method/design to arrive at valid findings, comparisons and conclusions
- This step explains how you will find answers to your research questions
- Must be valid, workable, and manageable
- Need strong reasons for selecting a particular design
  - Able to justify your selection
  - Aware of its strengths, weaknesses and limitations

Quantitative Research Approaches



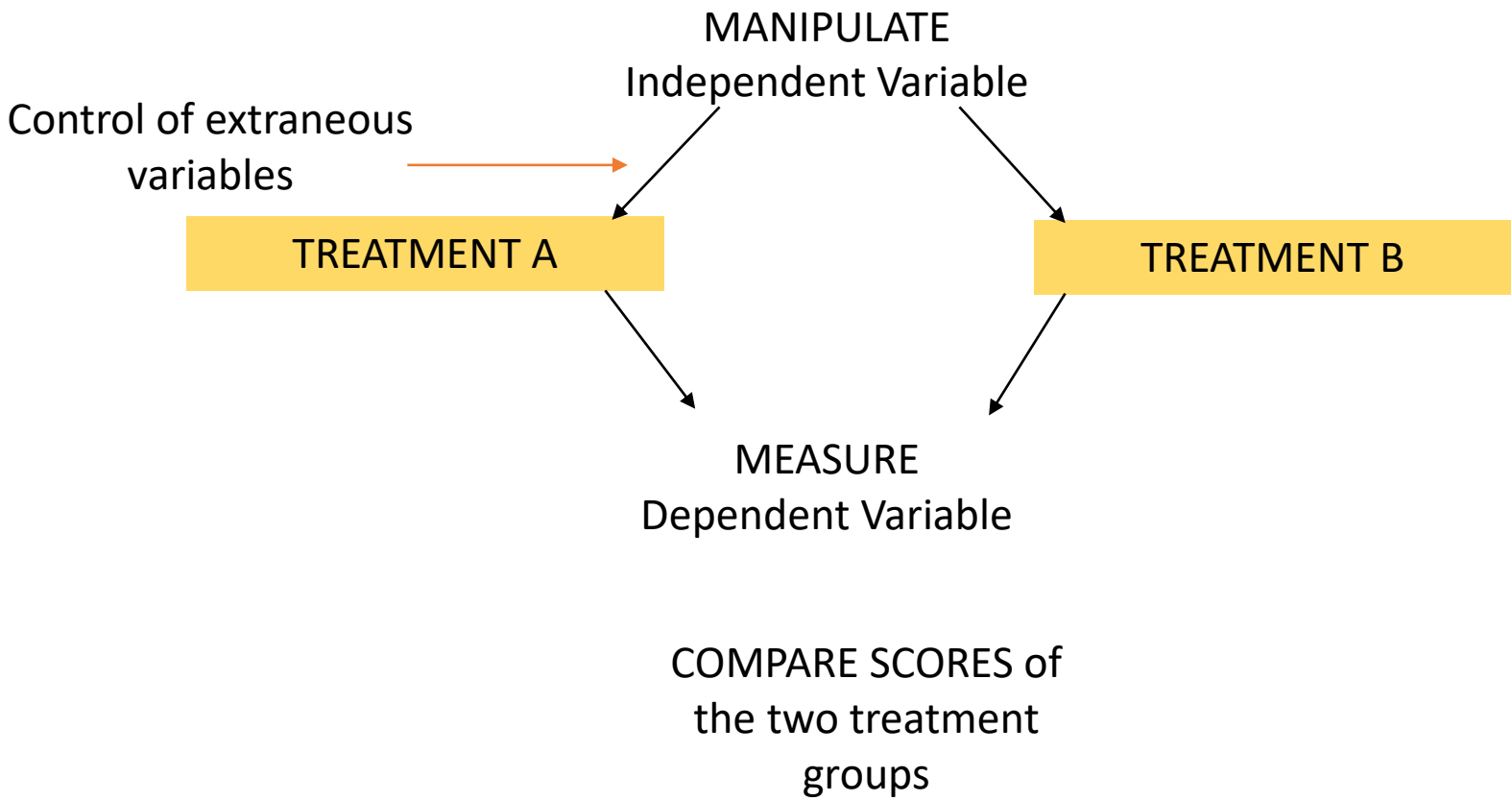
Quantitative Research Design



Experimental research

- Involve study of the effect of the systematic **manipulation** of one variable(s) on another variable.
- The manipulated variable is called the **experimental treatment** or the **independent variable.**
- The observed and measured variable is called the **dependent variable.**
- The final component of an experimental study is to control **extraneous variables** .
  - An extraneous variable is any variable other than the independent variable that might influence the dependent variable.
- In the manipulation, the researcher systematically introduces change (an intervention) and then observes the consequences of that change
- Experimental research involves the testing of hypotheses about the effect of one variable on another.
- The goal of experimental research is to determine whether a **causal relationship** exists between two or more variables

Basic Elements of An Experiment



Types of Experimental Research Design

- Classification according to number of independent variable
  - Single-variable design has one manipulated independent variable
  - Factorial designs have two or more independent variables, at least one of which is manipulated
- Classification according to control over threats to internal validity

(**Internal validity** is the extent to which we are confident that a cause-and-effect relationship established in a study cannot be explained by other factors.)

  - Pre-experimental - do not have random assignment of subjects to groups or other strategies to control extraneous variables
  - True-experimental - (also called randomized designs) use randomization and provide maximum control of extraneous variables
  - Quasi-experimental - lack randomization but employ other strategies to provide some control over extraneous variables

Pre-experimental design

- Provide little or no control of extraneous variables
- A weak designs
- Design 1: One-group pretest-posttest design
  - involves three steps:
    1. administering a pretest measuring the dependent variable;
    2. applying experimental treatment *X* to the subjects; and
    3. administering a posttest, again measuring the dependent variable.
- Design 2: Static group comparison
  - Uses two or more preexisting or intact (static) groups, only one of which is exposed to the experimental treatment.
  - Although this design uses two groups for comparison, it is flawed because the subjects are not randomly assigned to the groups and no pretest is used.
  - The researcher makes the assumption that the groups are equivalent in all relevant aspects before the study begins and that they differ only in their exposure to the treatment.

True-experimental design

- The designs in this category are called *true experiments* because subjects are randomly assigned to an experimental group and at least one control group.
- Because of the control they provide, they are the most highly recommended designs for experimentation ineducation.
- Example:
  - Design 3: Randomized Subjects, Posttest-Only Control Group Design
  - Design 4: Randomized Matched Subjects, Posttest-Only Control Group Design
  - Design 5: Randomized Subjects, Pretest–Posttest Control Group Design
  - Design 6: Solomon Three-Group Design
  - Design 7: Solomon Four-Group Design

Factorial design

- In complex social phenomena several variables often interact simultaneously, and restricting a study to one independent variable may impose an artificial simplicity on a complex situation.
- Involves manipulation of more variables simultaneously in order to study the independent effect of each variable on the dependent - variable, as well as the effects caused by interactions among the several variables.
- Example
  - Design 8: Simple factorial design

Quasi-experimental design

- Sometimes, it is not possible to randomly assign subjects to treatment groups
- **Quasi-experimental designs** are similar to randomized experimental designs in - that they involve manipulation of an independent variable but differ in that subjects are not randomly assigned to treatment groups.
- Because quasi-experimental designs do not provide full control, it is extremely - important that researchers be aware of the threats to both internal and external validity and consider these factors in their interpretation.
- Example
  - Design 9: Nonrandomized Control Group, Pretest–Posttest Design
  - Design 10: Counterbalanced Design

Designs		
Pre-experimental	**True-experimental**	**Quasi-experimental**
One-group pretest–posttest	Randomized Ss, posttest-only control group	Nonrandomized control group, pretest–posttest
Static group comparison	Randomized matched Ss, posttest-only control group	Counterbalanced
	Randomized Ss, pretest–posttest control group	One-group time series
	Solomon, three groups	Control group time series
	Solomon, four groups	
	Simple factorial	

Non-experimental research

- **Researcher identifies variables** and may look for relationships among them but **does not manipulate** the variables
- Survey research (also called descriptive research)
  - uses instruments such as questionnaires and interviews to gather information from groups of individuals.
  - Surveys permit the researcher to summarize the characteristics of different groups or to measure their attitudes and opinions toward some issue.
- Correlational research
  - gathers data from individuals on two or more variables and then seeks to determine if the variables are related (correlated).
  - Correlation means the extent to which the two variables vary directly (positive correlation) or inversely (negative correlation).
  - The degree of relationship is expressed as a numeric index called the coefficient of correlation.
- Ex-post facto/causal comparative
  - similar to an experiment, except the researcher does not manipulate the independent variable, which has already occurred in the natural course of events

Census Vs Sample Survey

- A survey that covers the entire population of interest is referred to as a **census**
- A survey that studies only a portion of the population is known as a **sample survey**

Cross-sectional study

- A cross section (sample) of a population at a single point in time.
  - In a longitudinal study of vocabulary development, for example, a researcher would compare a measure of first-grade students' vocabulary skills in 2000 with one when they were fourth-grade students in 2003 and seventh-grade students in 2006.
  - A cross-sectional study would compare the vocabulary skills of a sample of children from grades 1, 4, and 7 in 2006.
- The cross-sectional survey is the method of choice if you want to gather the data at one point in time.

Longitudinal survey

- Longitudinal surveys gather information at different points in time in order to study the changes over extended periods of time.
- Three different designs are used in longitudinal survey research: panel studies, trend studies, and cohort research.
- **Panel Studies** - the same subjects are surveyed several times over an extended period of time.
  - For example, a researcher studying the development of quantitative reasoning in elementary school children would select a sample of first-graders and administer a measure of quantitative reasoning. This same group would be followed through successive grade levels and tested each year to assess how quantitative reasoning skills develop over time.
  - Because the same subjects are studied over time, researchers can see the changes in the individuals' behavior and investigate the reasons for the changes.

…cont.

- **Trend Studies** -differs from a panel study in that different individuals randomly drawn from the same general population are surveyed at intervals over a period of time.
  - For example, researchers who have studied national trends in mathematics achievement sample middle school students at various intervals and measure their math performance.
  - Although the same individuals are not tested each time, if the samples from the population of middle school students are selected randomly, the results each time can be considered representative of the middle school population from which the student samples were drawn.
  - Test scores from year to year are compared to determine if any trends are evident.
- Another example of a trend study is the survey on alcohol, tobacco, and other drug use among Indiana youth conducted annually since 1991 by the Indiana Prevention Resource Center.
  - The 2007 survey collected data from 158,632 pupils in public and private schools in Indiana.
  - Results include a continuing long-term decline in cigarette smoking, an increase in binge drinking, a 3-year decline in marijuana use, and no change in the use of injection drugs.
  - Among the relationships found was that adolescents involved in after-school activities were less likely to use alcohol, tobacco, and drugs.
- This continuing trend study permits researchers to evaluate the prevention and enforcement efforts directed at the teenage population and to plan future programs.

…cont.

- **Cohort Studies** - a specific population is followed over a length of time with different random samples studied at various points.
- Whereas trend studies sample a general population that changes in membership over time, a cohort study samples a specific population whose members do not change over the duration of the survey.
- Typically, a cohort group has age in common.
  - For example, a school system might follow the high school graduating class(es) of 2004 over time and ask them questions about higher education, work experiences, attitudes, and so on.
  - From a list of all the graduates, a random sample is drawn at different points in time, and data are collected from that sample.
- Thus, the population remains the same during the study, but the individuals surveyed are different each time.

Ex-Post facto

- Ex-post facto/causal comparative
  - similar to an experiment, except the researcher does not manipulate the independent variable, which has already occurred in the natural course of events
- This method is sometimes called causal comparative because its purpose is to investigate cause-and-effect relationships between independent and dependent variables.
- Researchers use it in situations that do not permit the randomization and manipulation of variables characteristic of experimental research.
- Thus, much of the basic rationale for experimental and ex post facto is the same. They both investigate relationships among variables and test hypotheses.
- Ex post facto research is also appropriate when the variable actually could be manipulated but is not because it would be unethical or irresponsible to do so.
  - For example, it would not be ethical to manipulate illegal drug use or use of alcohol or cigarettes to study their effects on human subjects.

Comparing Different Designs

True Experimental	Qusai Experimental	Ex Post Facto

• Randomization
• Strict control of independent variable
• Large N
• Benefits
  • Imply cause & effect
  • Internal validity ↑
  • Greater degree of control and as a result a greater
• Drawback(s)
  • Time
  • Budget
• Stats (E.g.)
  • ANOVA
  • T-test
	• No Randomization
	  • *Randomness is neither possible nor practical*
	• Associate w/ blocked/attribute independent variable
	• Control group and/or multiple companions
	  • It's similar as experimental design except it makes use of naturally occurring groups rather than random assigning subjects to groups.
	• Benefits
	  • Effort to max External Validity
	  • Natural settings (War, disaster, winning lottery)
	• Drawbacks
	  • No R → generalize ↓
	  • Ability to use inferential stats ↓
	• Stats (E.g.)
	  • ANOVA
	  • T-test
		• No Randomization
		• No control over independent variable
		• Benefits
		  • External validity ↑
		  • Cheap, fast, convenient
		  • Medicine uses this method widely
		• Drawbacks
		  • No cause & effect
		  • Internal Validity ↓
		• Stats (E.g.)
		  • T-test
		  • ANOVA
		• *Similar: correlational*
		  • *Involves looking at existing circumstances*

Correlational research

- The most useful applications of correlation are (1) assessing relationships, (2) assessing consistency, and (3) prediction.
- Does not determine cause-and- effect relationship
- Correlational research
  - gathers data from individuals on two or more variables and then seeks to determine if the variables are related (correlated).
  - Correlation means the extent to which the two variables vary directly (positive correlation) or inversely (negative correlation).
  - The degree of relationship is expressed as a numeric index called the coefficient of correlation.

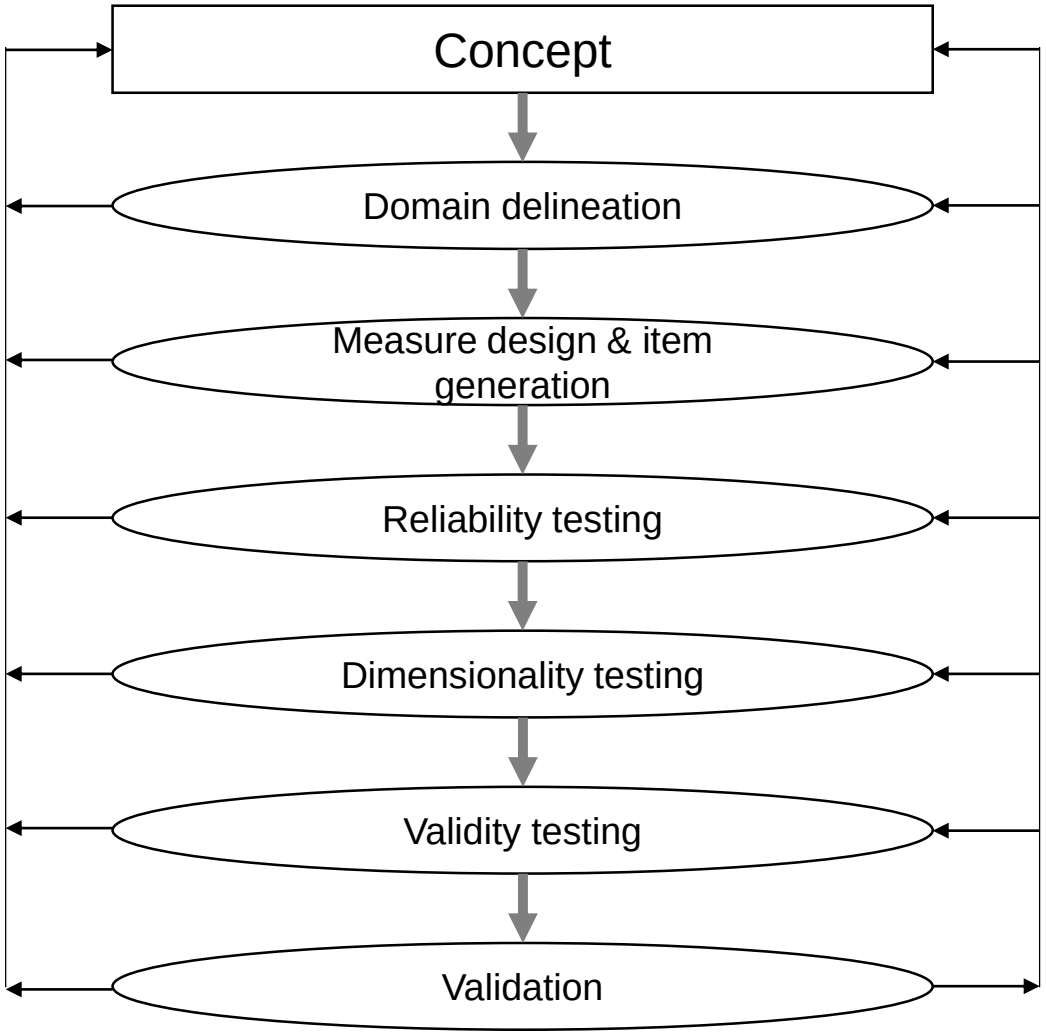
Statistical Analysis according to types of Research Design

Research Design	Statistical Analysis Applicable
Experimental designs	T-tests, One-Way ANOVA, Factorial ANOVA
Longitudinal studies	One-Way ANOVA, Factorial ANOVA
Cross sectional	T-tests, One-Way ANOVA, Factorial ANOVA, correlations, regressions
Descriptive	Descriptive analysis (MCT, MD, MP)
Ex-post Facto	T-tests, One-Way ANOVA
Correlational	Correlations, Regression, chi-square for relationships between categorical variables

Step 3: Constructing an Instrument for Data Collection

- Research tool/research instrument - Anything that becomes a means of collecting information for your study
  - Observation forms, interview schedules, questionnaires and interview guides
- Pilot-testing a research tool is an integral part of instrument construction.
  - As a rule, the pilot test of a research instrument should not be carried out on the sample of your study population but on a **similar** population which you are not proposing to study.

Steps in Measure Development Process



Viswanathan (2015, p.6)

Adopting a Measurement

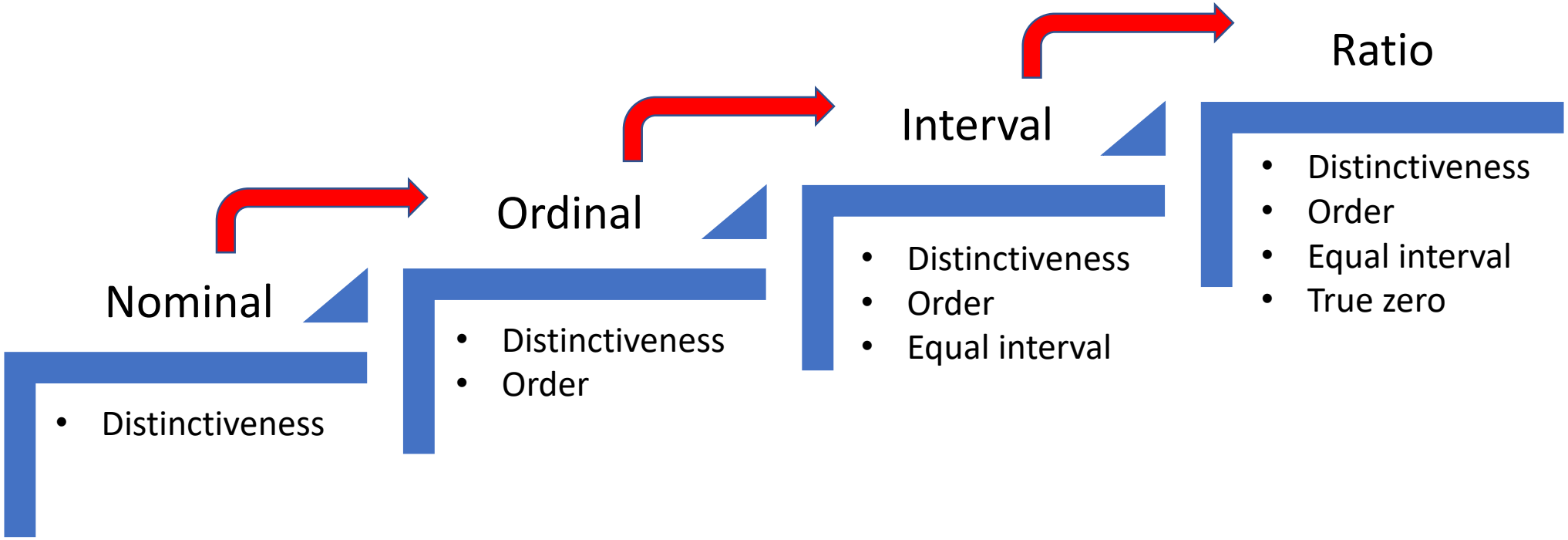
- Adopt – no changes should be made in wording, answer categories, order or format of questions, or administrative procedures

- Adopt when
  1. Replicate a study
  2. Incapable to improve another's instrument
  3. Instrument is under copyright
  4. To compare the findings
  5. To use instruments of known reliability & validity

Adapting a Measurement

- Adapt – any aspect of and instrument is changed
- Adapt when
  1. Instrument is too long
  2. Different population from the original population is being studies
  3. Instrument need to be translated to other language
  4. Instrument need to be expanded, reordered
  5. Change in the data collection procedure
- Pilot test should be repeated & instrument's reliability & validity must be revaluated

Scale of Measurement



What Statistics are Appropriate?

Scale of Measurement	Type of appropriate statistics
Nominal	• Frequency, percentage, Measure of Central Tendency (mode), chi-square correlation (contingency correlation).
Ordinal	• All calculation of the previous level, nominal, might be performed. • Median might be calculated as a measure of central tendency. Percentiles, Quartiles and some additional nonparametric statistics might be used. • i.e., Spearman's rank-order correlation might be used to calculate the correlation between two variables measured at the ordinal level.
Interval	• Almost every statistical tool available. • All the previously allowed tools might be used • Mean and standard deviation, both frequently used calculations of central tendency and variability, become available for use at this level. • The only statistical tools that should not be used at this level are those that require the use of ratios, such as the coefficient of variation.
Ratio	• All statistical tools are available for data at this level

Validity Overview

- A study is valid if its measures actually measure what they claim to
- The degree to which a measure is free of random and systematic error (Nunnally, 1978)
- Types of validity
  - Face validity
  - Content validity
  - Criterion validity
  - Construct validity

- Test validity when
  - Develop new instrument
  - Use established instruments to new populations

Validity by type, use, and method to test and strength of estimate

Type	Use	Method to test	Strength of estimate
Face validity	Is the measure testing what it is supposed to measure? Are the items plausible?	Peer review	No statistical inference testing
Content validity	Does the measurement instrument reflect a specific domain of content?	Content validity index (CVI)	$CVI \geq 80\%$
Criterion validity	Can the outcome of one assessment be used as a substitute test to the established *gold-standard criterion* test?	Prediction methods	See below
• Concurrent validity	Do the results of a criterion measure and a target test, given at the same relative time point, concur with one another?	Receiver operating characteristics (ROC) curves	Area under the curve (AUC) $\geq .70$ with $p \leq .05$
• Predictive validity	Does the outcome of a target test predict a future criterion score or outcome?	Prediction methods, i.e., regression analysis	The higher the R^2in the prediction model, the better the validity
Construct validity	Does the assessment measure a construct and the theoretical components underlying the construct?	Test of correlation	Higher correlation coefficient are better, thus $\mathrm{r} \geq .80$
• Convergent	Is the level of agreement between two tests that are being used to measure the same construct acceptable?	Test of correlation	Higher correlation coefficient are better, thus $\mathrm{r} \geq .80$
• Discriminant validity	Is the level of disagreement (poor or zero correlation) when two tests measure a trait, behavior, or characteristic acceptable?	Test of correlation	Higher correlation coefficient are better, thus $\mathrm{r} \geq .80$

Reliability

- Reliability refers to the consistency of scores, that is, an instrument's ability to produce "approximately" the same score for an individual over repeated testing or across different raters (Lodico, Spaulding, & Voegtle, 2010, p. 93)
  - A reliable instrument can generate consistent results.
  - When an instrument is applied to target subjects more than once, an investigator can expect to obtain results that are quite similar or even identical each time.
- Such measurement consistency enables investigators to gain confidence in measuring ability or dependability of the particular instrument.
- Reliability is a concept that is expressed in terms of correlations that are summarized in a reliability coefficient.
  - reliability coefficient can assume values from zero to +1.00. The closer to +1.00 the reliability coefficient is, the more highly reliable the instrument.
- Approaches to reliability consist of repeated measurements on an individual (i.e., test–retest and equivalent forms), internal consistency measures (i.e., split-half, Kuder– Richardson 20, Kuder–Richardson 21, and Cronbach's alpha), and interrater and intrarater reliability.
- Usually, reliability is shown in the numerical form, as a coefficient.
- The range of reliability coefficient is from 0 (errors existed in the entire measurement) to 1 (no error in the measurement was discovered); the higher the coefficient, the better the reliability.

Reliability

- Reliability relates only to consistency
  - A reliable person who is consistent but always 15 minutes late would still be reliable in a measurement context
- A measurement can be reliable but not necessary accurate (valid)
- Types of reliability
  - Test-retest reliability
  - Alternate form reliability
  - Internal consistency reliability

Factors Influencing Reliability

- Test taker's personal characteristics (for example, motivation health, mood, fatigue)
- Variations in test setting (for example, differences in the physical characteristics of the room)
- Variations in the administration and scoring of the test
- Variation in participant responses due to guessing
- The instrument (Ary et al., 2010)
  - ☞ Luck is more of a factor in a short test than in a long test.

(Lodico, Spaulding, & Voegtle, 2010, p. 94)

PILOT STUDY/TEST

- After the questionnaire measuring all variables has been constructed, it is necessary to pilot the form on a representative sample of about 10–30 respondents.
- Pilot study involves the administration of the survey to a small group of individuals that share similar characteristics with actual research respondents
- Provides the opportunity to assess the appropriateness of the data-collection methods and other procedures and to make changes if necessary.
- It also permits a preliminary testing of the hypothesis, which may give some indication of its tenability and suggest whether further refinement is needed.
- This process helps the researcher determine the survey's validity and its reliability.
- Clear instructions and clear understanding of the process are necessary to achieve data of good quality.

Step 4: Selecting a Sample

- Two key aims
  - Avoid bias
  - Maximum precision from sample to population
- Types of sampling designs
  - random/probability sampling designs
  - non-random/non-probability sampling designs (purposive sampling)
  - ‘mixed’ sampling design
- Type of sampling strategy used will influence your ability to make generalizations from the sample findings about the study population, and the type of statistical tests you can apply to the data.

Unit of Analysis

- **Sampling frame** defines the population from which the sample will be drawn and to which the sample data will be generalized (Nicholas, 2009, p.31)
- Unit of analysis is the major entity that you are analysing in your study
  - It is the analysis you do in your study that determines what the unit is
  - i.e.,
    - If you are comparing the children in two classrooms on achievement test scores, the unit is the individual child because you have a score for each child.
    - If you are comparing the two classes on classroom climate, your unit of analysis is the group, in this case the classroom

Sampling

- Process through which a sample is extracted from a population
- The more the sample is representative of the population, the higher is the accuracy of the inferences and better are the results generalisable.
- A sample is said to be representative when the characteristics of elements selected are similar to that of entire target population.
- The results are said to be generalisable when the findings obtained from sample are equally true for the entire target population.
- Sampling process may encounter the problem of systematic errors and sampling biases.
  - Systematic errors can be defined as incorrect or false representation of the sample caused by over representation of one characteristic and/or under representation of the others.
  - Sampling bias is said to occur when the selected sample does not truly reflect the characteristics of population.

Key Factors in Sampling

- Sample size
- Representativeness and parameters of the sample
- Access to the sample
- Sampling strategy to be used

Cohen, Manion, & Morrison (2007, pg. 100)

Sample Size

- No clear-cut answer as to what constitutes an adequate, or sufficient, size for a sample.
- Continuous data require smaller sample size, Categorical data requires larger sample size (Bartlett, Kotrlik, & Higgins, 2001)
- A recommended minimum number of subjects is:
  - 100 for a descriptive study
  - 50 for a correlational study,
  - 30 in each group for experimental and causal-comparative studies

(Fraenkel, Wallen, & Hyun, 2008, p.107).

Sample Size – Continuous Data

Cochran's sample size formula

$$n_0 = \frac{(t)^2 * (s)^2}{(d)^2}$$

Where,
t = value for selected alpha level of .025 in each tail = 1.96 (the alpha level of .05 indicates the level of risk the researcher is willing to take that true margin of error may exceed the acceptable margin of error.)
s = estimate of standard deviation in the population = 1.167.
(estimate of variance deviation for 7 point scale calculated by using 7 [inclusive range of scale] divided by 6 [number of standard deviations that include almost all (approximately 98%) of the possible values in the range]).
d = acceptable margin of error for mean being estimated = .21.
(number of points on primary scale * acceptable margin of error; points on primary scale = 7; acceptable margin of error = .03 [error researcher is willing to except]).

Example

- Assume that a researcher has set the
  - $\alpha = .05$, plans to use a seven-point scale,
  - Error = 3%
  - SD = 1.167.

$$n_0 = \frac{(t)^2 * (s)^2}{(d)^2}$$

$$n_0 = \frac{(1.96)^2 * (1.167)^2}{(7 * .03)^2}$$

$$n_0 = 118$$

Therefore, for a population of 1,679, the required sample size is 118.

…cont.

- However, since this sample size exceeds 5% of the population (1,679*.05=84), Cochran's (1977) correction formula should be used to calculate the final sample size. These calculations are as follows

$$n = \frac{n_0}{(1 + \frac{n_0}{population})}$$

$$n = \frac{118}{(1 + \frac{118}{1679})}$$

$$n = 111$$

…cont.

- These procedures result in the minimum returned sample size.
- If a researcher has a captive audience, this sample size may be attained easily. However, since many educational and social research studies often use data collection methods such as surveys and other voluntary participation methods, the response rates are typically well below 100%.
- Salkind (1997) recommended oversampling when he stated that “If you are mailing out surveys or questionnaires, ... count on increasing your sample size by 40%-50% to account for lost mail and uncooperative subjects” (p. 107).
- Fink (1995) stated that “Oversampling can add costs to the survey but is often necessary” (p. 36).
- Cochran (1977) stated that “A second consequence is, of course, that the variances of estimates are increased because the sample actually obtained is smaller than the target sample.
- This factor can be allowed for, at least approximately, in selecting the size of the sample” (p. 396).
- However, many researchers criticize the use of over-sampling to ensure that this minimum sample size is achieved and suggestions on how to secure the minimal sample size are scarce.

…cont.

- If the researcher decides to use oversampling, four methods may be used to determine the anticipated response rate:
  - (1) take the sample in two steps, and use the results of the first step to estimate how many additional responses may be expected from the second step;
  - (2) use pilot study results;
  - (3) use responses rates from previous studies of the same or a similar population; or
  - (4) estimate the response rate.
  - The first three ways are logical and will produce valid estimates of response rates; therefore, they do not need to be discussed further. Estimating response rates is not an exact science. A researcher may be able to consult other researchers or review the research literature in similar fields to determine the response rates that have been achieved with similar and, if necessary, dissimilar populations.
- Therefore, in this example, it was anticipated that a response rate of 65% would be achieved based on prior research experience. Given a required minimum sample size (corrected) of 111, the following calculations were used to determine the drawn sample size required to produce the minimum sample size:
- Where anticipated return rate = 65%.
  Where n_2= sample size adjusted for response rate.
- Where minimum sample size (corrected) = 111.
- Therefore, n_2 = 111/.65 = 171.

Sample Size – Categorical Data

- Assume a researcher has set
  - $\alpha = .05$, plans to use a proportional variable,
  - level of acceptable error at 5%,
  - SD =.5

$$n_0 = \frac{(t)^2 * (p)(q)}{(d)^2}$$

$$n_0 = \frac{(1.96)^2 * (.5)(.5)}{(.05)^2}$$

Therefore, for a population of 1,679, the required sample size is 384

Where t = value for selected alpha level of .025 in each tail = 1.96.
(the alpha level of .05 indicates the level of risk the researcher is willing to take that true margin of error may exceed the acceptable margin of error).
Where (p)(q) = estimate of variance = .25.
(maximum possible proportion (.5) * 1-maximum possible proportion (.5) produces maximum possible sample size).
Where d = acceptable margin of error for proportion being estimated = .05 (error researcher is willing to except).

…cont.

- However, since this sample size exceeds 5% of the population (1,679*.05=84), Cochran's (1977) correction formula should be used to calculate the final sample size. These calculations are as follows:

$$n = \frac{n_0}{(1 + \frac{n_0}{population})}$$

$$n = \frac{384}{(1 + \frac{384}{1679})}$$

$$n = 313$$

Where population size = 1,679
Where n_0 = required return sample size according
to Cochran's formula= 384
Where n_1 = required return sample size because sample > 5% of population

…cont.

- These procedures result in a minimum returned sample size of 313.
- Using the same oversampling procedures as cited in the continuous data example, and again assuming a response rate of 65%, a minimum drawn sample size of 482 should be used.
- These calculations were based on the following:
  - Where anticipated return rate = 65%.
  - Where n_2= sample size adjusted for response rate.
  - Where minimum sample size (corrected) = 313.
  - Therefore, n_2 = 313/.65 = 482.

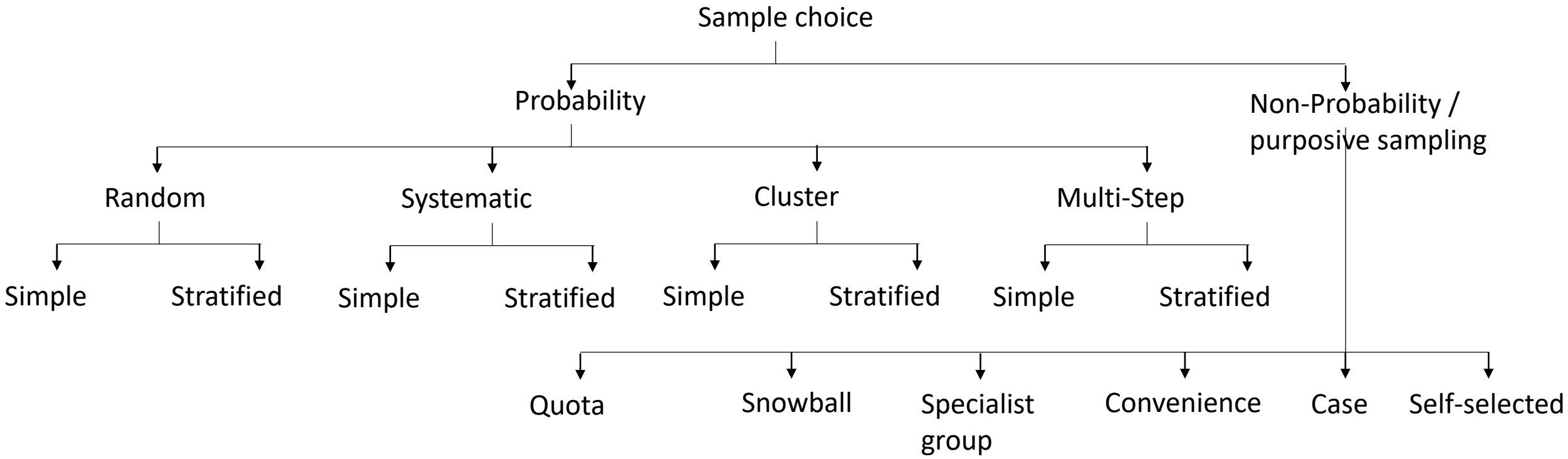
Online Sample Size Calculator

- Raosoft
- Qualtrics
- G*Power

Representative Samples

- Finding samples that are representative of the population from which they are drawn, so that generalisations can be made to the rest of it.
- 'Representative' implies that
  - the sample accurately reflects the composition of the population from which it is drawn;
  - it has the same characteristics (apart from its size) in the same proportion.
- Able to generalise results to the whole population if sample is representative
  - Need to have an unbiased sample of the population (representative sample means the selection of sample do not skewed towards one group or another)
- Need to reduce bias to ensure sample representativeness

Sampling Procedures



Newby, 2014, p.238

Types of Sampling

- Probability
  - Elements are drawn by chance procedures
  - every element in the population has an equal chance of being selected
  - Types of probability sampling
    - Simple random sampling
    - Stratified random sampling
    - Cluster sampling
- Non-probability
  - Elements are not chosen by chance procedures
  - Success depends on the knowledge, expertise, and judgment of the researcher.

Simple Random Sampling

- Steps
  1. Define the population.
  2. List all members of the population.
     - assign each member of the population a distinct identification number
  3. Select the sample by employing a procedure where sheer chance determines which members on the list are drawn for the sample) i.e, lottery method)
     - Write the student numbers on separate slips of paper
     - Place the pieces of paper in a container
     - Shake the container, and draw out a slip of paper
     - Shake the container again, draw out another paper, and continue the process until reach the targeted sample size.
- Random number generator online (i.e., Research Randomizer; Random Lists)
- Table of random numbers

Summarizing Simple Random Sampling

- The size of the population must be known and the elements that make up the population (schools, students, teachers, etc.) must be identifiable.
- Random selection means that every element of the population must have the **same** chance of being selected.
- The process generates data that is appropriate for advanced statistical analysis.
- It can be costly to collect data if the elements in the sample are widely spaced and we have to travel to the person or place selected in order to collect our data.

Stratified Random Sampling

- Works by separating the population into categories
- Advantage: increases the likelihood of representativeness, especially if one's sample is not very large.
- Disadvantage: requires more effort on the part of the researcher.

Example

Schools

Business studies	Sports	Science	Vocational
18%	27%	36%	19%

If n=280,




(280 x 18% = 50)	(280 x 27% = 76)	(280 x 36% = 101)	(280 x 19% = 53)

Summarizing Stratified Random Sampling

- The groups (strata) that are used must be meaningful in terms of the research issue or analysis.
- The relative size of each group in relation to the total population must be known.
- The sample size of a stratum is determined by the stratum's proportion of the total population.
- Samples are chosen by a random or systematic procedure.
- Overall sample size may be reduced by treating each stratum as a population and determining the sample size for each stratum separately.

Cluster Random Sampling

- Random selection of groups, or clusters, of subjects rather than individuals
- Applicable when it is not possible to select a sample of individuals from a population
  - Sometimes, for example, a list of all members of the population of interest is not available.

Example

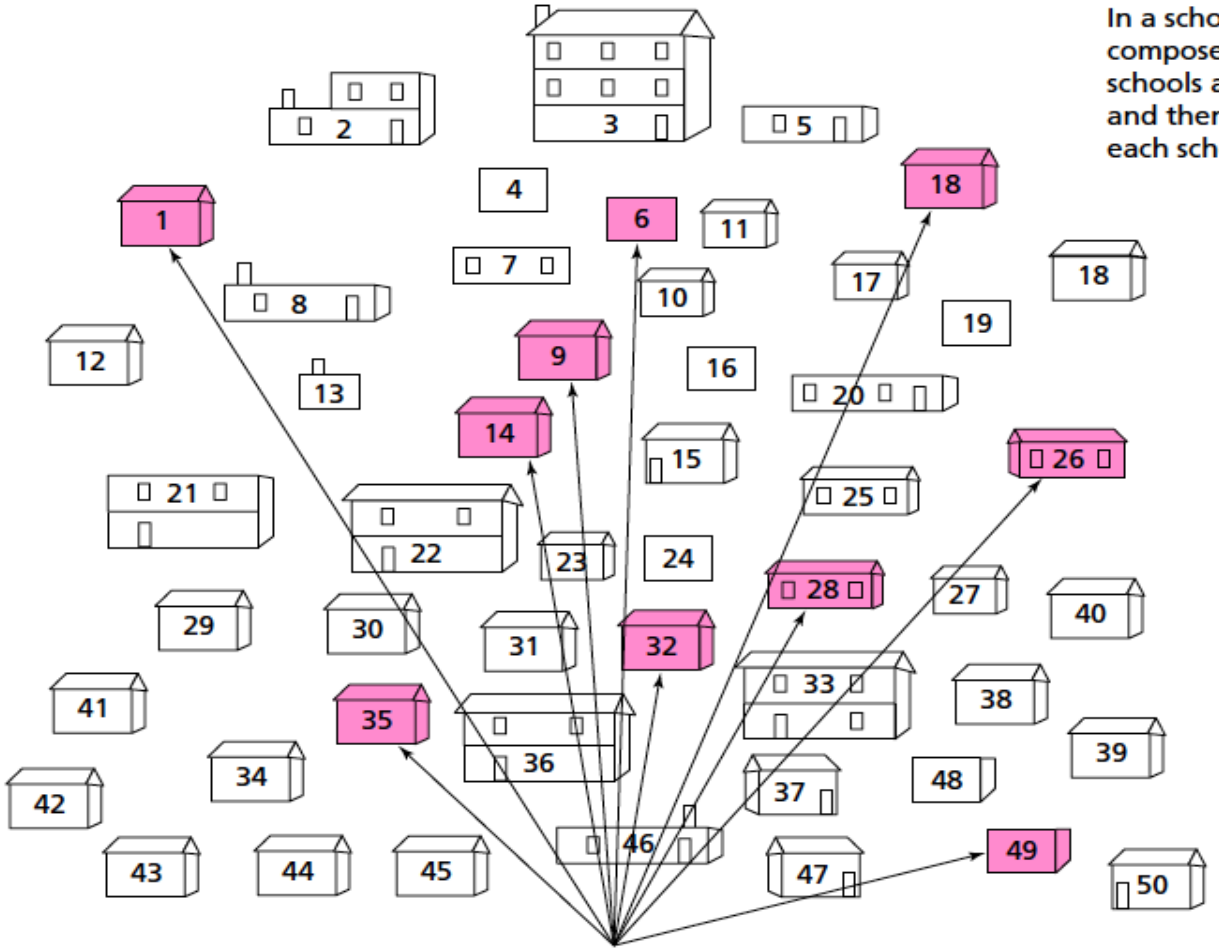


Figure 6.3 *Cluster Random Sampling*

Summarizing Cluster Random Sampling

- Cluster sampling is not as effective at representing the true characteristics of a population as random or systematic sampling.
- Cluster sampling has major advantages in terms of time and cost.
- Clusters should be diverse in character.
- Clusters and respondents are selected by random or systematic sampling procedures.

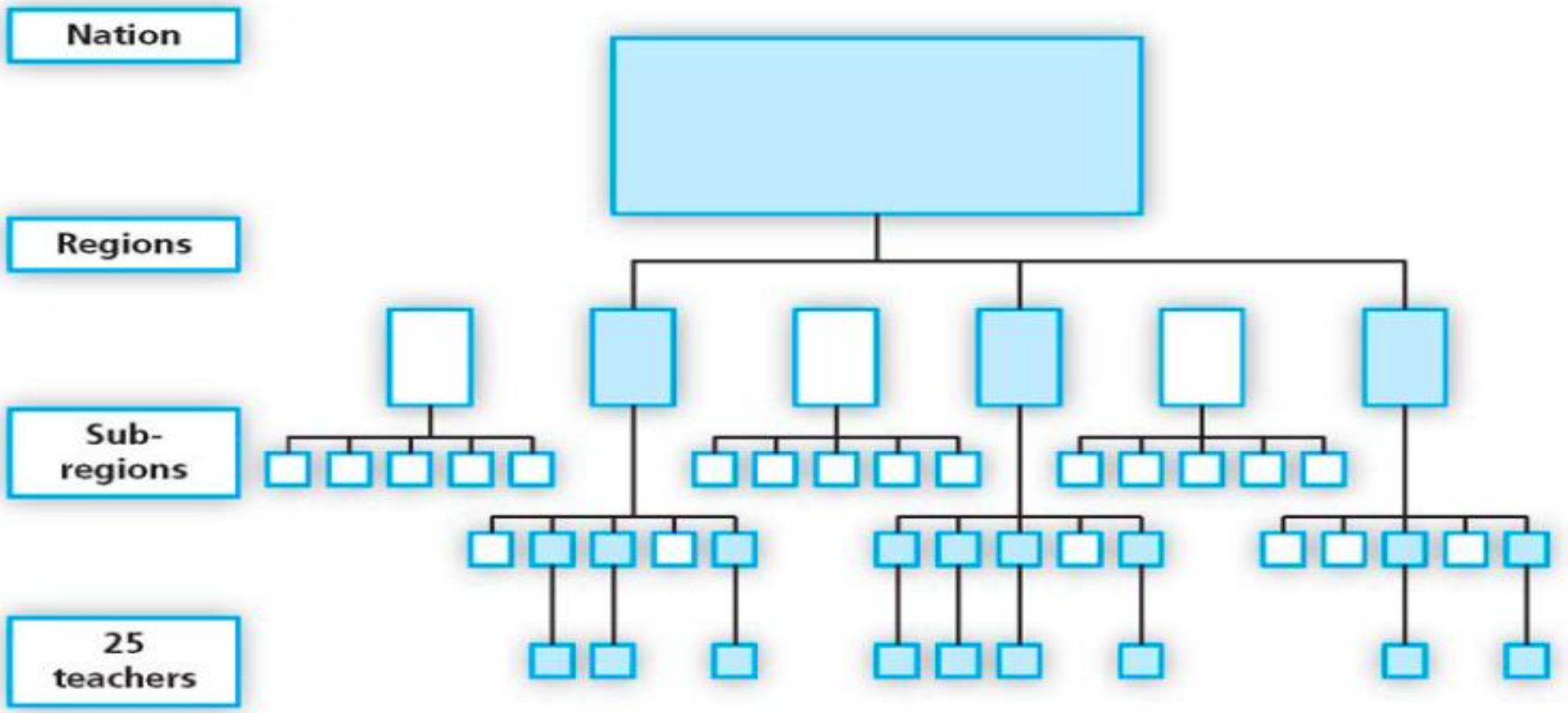
Systematic Random Sampling

- Unlike simple random sampling, there is not an equal probability of every element been included.
- In this type of sampling the elements are selected at a regular interval.
- The interval may be in terms of time, space or order. For instance, element appearing after every 30 minutes, or present at a distance of two meters, or every 5th element present on a list.
- This regularity and uniformity in selection makes the sampling systematic.
- List of elements may or may not be required before the conduction of research.
- Sometimes it is not even possible to create a list because of the nature of population.
  - Say, if it is possible to tell who is going to visit the coffee shop today.

Summarizing Systematic Random Sampling

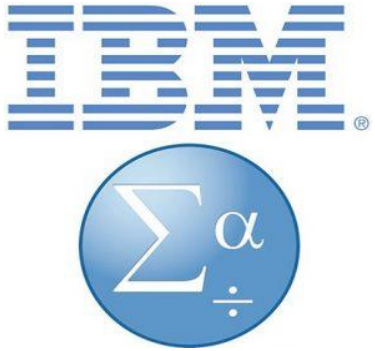
- The size and sometimes the spatial distribution of the target population must be known in order to devise the sample frame.
- After the first data source has been identified randomly, the systematic procedure of sampling 1 in n generates results that are the equivalent of a random sample.
- Statistical tests can be used on data generated by a systematic sample.
- Use of a systematic sample can generate significant savings in time and cost in comparison with the equivalent random sample.

Multi-stage/ Multi-Phase Random Sampling



Diagrammatic representation of multi-phase sampling

Step 7: Processing and Displaying Quantitative Data



Minitab® 18

STATISTICA®

Parametric Vs Non-Parametric Data

- Non-parametric data – Data that does not met the assumption of normality
- Parametric data – Data that met the assumption of normality
  - Nominal and ordinal data considered to be non-parametric
  - Interval and ratio data considered to be parametric data (unless, the data are skewed).
- It is incorrect to apply parametric statistics to non-parametric data, but it is possible to apply non-parametric statistics to parametric data if those data do not conform to the curve of distribution, being skewed or unevenly distributed.
- Statistics for parametric data tend to be more powerful than those for non-parametric data, though such power is bought at the price of, for example, conformity to the normal curve of distribution and random samples.

With Knowledge We Serve

Descriptive Statistics

- Defined by what descriptive statistics are not: not inferential, not multivariate, or not causal.
- Describe and present data, for example, in terms of summary frequencies.
- No attempt to infer or predict population parameters
- This includes:
  - Measure of Central tendency
    - mean (the average score);
    - mode (the score obtained by the greatest number of people);
    - median (the score obtained by the middle person in a ranked group of people, i.e. it has an equal number of scores above it and below it);
  - Measure of Dispersion
    - range (the distance between the highest and the lowest scores);
    - variance (a measure of how far scores are from the mean, calculated as the average of the squared deviations of individual scores from the mean);
    - standard deviation (a measure of the dispersal or range of scores, calculated as the square root of the variance, yielding the average of all the individual deviations of scores from the mean);

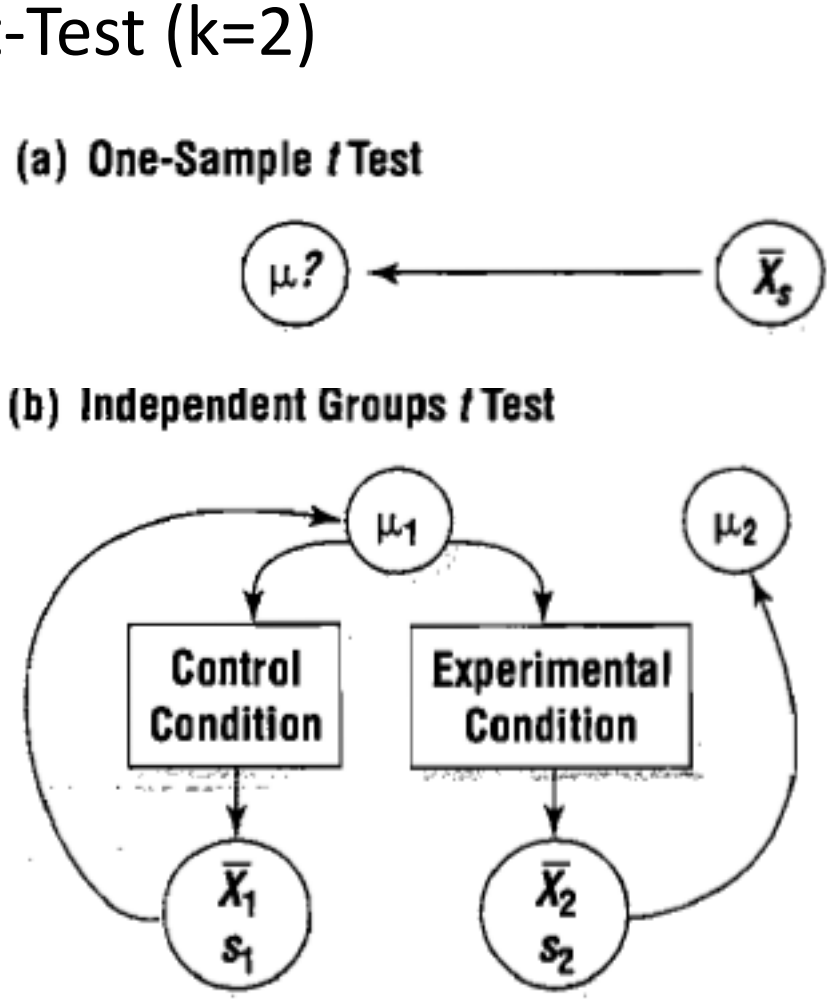
Inferential Statistics

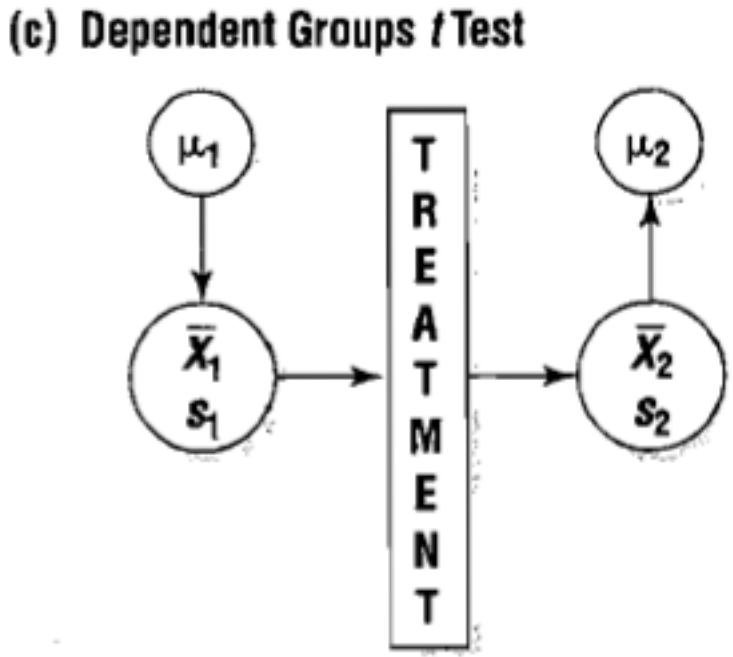
- Strive to make inferences and predictions based on the data gathered.
- They infer or predict population parameters or outcomes from sample statistics, based on probability
- Include hypothesis testing
  - Regression and multiple regression (highest level of analysis in statistics)
  - Correlation
  - Difference testing (e.g. t-tests and Analysis of Variance)
  - Factor analysis and structural equation modelling.
- More powerful

H_0: *the null hypothesis*
H_1: *the alternative hypothesis*

Types of Inferential Statistics (Comparison)

- t-Test (k=2)





…cont.

- One-Way ANOVA
  - A one-way ANOVA will compare the means of three or more levels with one another in order to determine if any significant difference(s) exist(s) between or among them.

Hypothesis for Differences

H_0: *Hypothesis of no difference*
H_1: *Hypothesis of the difference*

Example
H_0: *There is no statistically significant difference between males and females students in the results of Research Methodology*
H_1: *There is a statistically significant difference between males and females students in the results of Research Methodology*

Types of Inferential Statistics (Association)

- Pearson Product Moment Correlation
  - assess the nature of the association between two variables, X and Y
  - correlation does not imply causation
  - Described by index of association, Pearson *r*
    - a correlation coefficient, is a statistic that quantifies the extent to which two variables X and Y are associated, and whether the direction of their association is positive, negative, or zero.
    - *r* ranges from -1.0 to +1.0

Guildford Rule of Thumb

Value (Range)	Association (Strength)
< 0.2	Negligible Relationship
0.2 to 0.4	Low Relationship
0.4 to 0.7	Moderate Relationship
0.7 to 0.9	High Relationship
> 0.9	Very High Relationship

…cont.

- Spearman's Rank-Order Correlation (r_s)
  - Non-parametric test
  - calculates a correlation using ordinal data that have been ranked or interval data that does not met the assumption of normality
  - r_s ranges from -1.0 to +1.0

Guildford Rule of Thumb

Value (Range)	Association (Strength)
< 0.2	Negligible Relationship
0.2 to 0.4	Low Relationship
0.4 to 0.7	Moderate Relationship
0.7 to 0.9	High Relationship
> 0.9	Very High Relationship

Types of Inferential Statistics (Simple Linear Regression)

- Simple linear regression is an extension to Pearson Product Moment Correlation
  - it involves examining changes in the level of Y relative to changes in the level of X
- Purpose
  - Determine relationship between two variables, generally between IV and DV
  - Make prediction of DV based on IV $\boldsymbol{Y = b_0 + b_1 X_1}$

Hypothesis for Relationship

H_0: *Hypothesis of no relationship*
H_1: *Hypothesis of the relationship*

Example
H_0: *There is no statistically significant correlation between supervisor support and employee job performance*
H_1: *There is a statistically significant correlation between supervisor support and employee job performance*

UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI

Thank you

With Knowledge We Serve