

Pearson New International Edition

Research in Education

Evidence-Based Inquiry

James McMillan Sally Schumacher

Seventh Edition

Pearson New International Edition

Research in Education
Evidence-Based Inquiry
James McMillan Sally Schumacher
Seventh Edition

PEARSON

Pearson Education Limited

Edinburgh Gate

Harlow

Essex CM20 2JE

England and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsoned.co.uk

© Pearson Education Limited 2014

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a licence permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

PEARSON

ISBN 10: 1-292-02267-1

ISBN 13: 978-1-292-02267-3

British Library Cataloguing-in-Publication Data

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

Printed in the United States of America

Table of Contents

Glossary James H. McMillan/Sally Schumacher	1
1. Introduction to Evidence-Based Inquiry James H. McMillan/Sally Schumacher	9
2. Research Designs and Reading Research Articles James H. McMillan/Sally Schumacher	27
3. Research Problems, Questions, and Hypotheses James H. McMillan/Sally Schumacher	55
4. Review of Literature James H. McMillan/Sally Schumacher	83
5. Designing Quantitative Research: Purpose, Validity, and Ethical Considerations James H. McMillan/Sally Schumacher	113
6. Participants, Subjects, and Sampling for Quantitative Designs James H. McMillan/Sally Schumacher	141
7. Descriptive Statistics James H. McMillan/Sally Schumacher	161
8. Quantitative Data Collection: Technical Adequacy James H. McMillan/Sally Schumacher	187
9. Collecting Quantitative Data James H. McMillan/Sally Schumacher	203
10. Nonexperimental Research Designs, Surveys, and Secondary Data Analysis James H. McMillan/Sally Schumacher	233
11. Experimental, Quasi-Experimental, and Single-Subject Designs James H. McMillan/Sally Schumacher	275
12. Understanding and Reporting Inferential Data Analyses James H. McMillan/Sally Schumacher	315

13. Designing Qualitative Research James H. McMillan/Sally Schumacher	343
14. Collecting Qualitative Data James H. McMillan/Sally Schumacher	367
15. Qualitative Data Analysis and Narrative Structure James H. McMillan/Sally Schumacher	393
16. Mixed Method Designs James H. McMillan/Sally Schumacher	423
17. Concept Analysis and Historical Research James H. McMillan/Sally Schumacher	447
18. Evaluation Research and Policy Analysis James H. McMillan/Sally Schumacher	461
19. Action Research James H. McMillan/Sally Schumacher	477
Appendix: Guidelines for Writing Research Proposals James H. McMillan/Sally Schumacher	489
Appendix: Guidelines for Writing Research Reports James H. McMillan/Sally Schumacher	495
Appendix: Calculations for Selected Descriptive and Inferential Statistics James H. McMillan/Sally Schumacher	499
References James H. McMillan/Sally Schumacher	517
Index	521

Glossary

- A-B design** A single-subject design that compares frequency of behavior during the baseline (A) with intervention (B) conditions
- A-B-A design** A single-subject design that compares the baseline (A) with the intervention (B) and then with the baseline (A)
- abbreviated time-series design** Adaptation of the time-series design in which there are only a few pre- and post-measures prior to the intervention
- achievement tests** Tests that measure knowledge, skills, or behavior
- action research** Studies undertaken by practitioners in schools that address an actual problem or issue in the school or classroom
- agreement** A type of reliability based on the consistency of ratings or observations among two or more persons
- alpha level** Predetermined level of probability to reject the null hypothesis
- alternative assessment** Procedures used to measure performance through constructed-response answers, unlike traditional paper-and-pencil tests
- alternative hypothesis** A statistical statement that is opposite the null hypothesis
- analysis of covariance (ANCOVA)** An inferential statistical test used to adjust statistically the effect of a variable related to the dependent variable
- analysis of variance (ANOVA)** An inferential statistical procedure for determining the level of probability of rejecting the null hypothesis with two or more means
- analytical research** An analysis of documents to investigate historical concepts and events
- anonymity** Ensures that none of the participants are identified
- applied research** Research that is conducted in a field of common practice and is concerned with the application and development of research-based knowledge
- aptitude test** A test used to predict performance
- artifacts** Material objects of a current or past event, group, person, or organization that reveal social processes, meanings, and values
- assent** A procedure whereby children aged 7–17 agree or do not agree to participate in a study
- attenuation** The lowering of a measure of relationship between two variables because of the unreliability of the instruments used
- attrition** A threat to internal validity in which loss of subjects affects the results
- authenticity** The faithful reconstruction of participants' multiple perceptions
- basic research** Research that tests or refines theory; not designed to be applied immediately to practice
- behavioral objectives** Objectives of a practice that are stated in terms of observable terminal performances, which can be measured; also called *performance objectives* or *measured objectives*
- beta weight** A standardized regression coefficient
- bivariate** Refers to correlation between or testing of two variables or categories for differences
- bivariate regression** A regression analysis in which the dependent variable is predicted by a single independent variable
- boolean** Operators such as *and* and *or* that are used to limit a literature or Internet search
- box-and-whisker plot** A graphic illustration of variability of a set of scores
- case study** Qualitative research that examines a bounded system (i.e., a case) over time in detail, employing multiple sources of data found in the setting
- categorical variable** A variable used to divide subjects, objects, or entities into two or more groups
- categories** Abstract terms that represent the meaning of sets of related topics
- central question** A general question that identifies the main phenomenon that is examined in a qualitative study
- chi-square** A nonparametric statistical procedure that is used with nominal data to test relationships between the frequency of observations in categories of independent variables; also called *goodness of fit*
- closed form** A type of questionnaire item in which the subject chooses between or among predetermined options
- cluster sampling** A form of probability sampling in which subjects are first grouped according to naturally occurring traits
- code** A descriptive name for the subject or topic of a data segment
- coefficient of concordance** A type of interrater reliability based on the rank-order agreement among raters
- coefficient of determination** A squared correlation coefficient that indicates the percentage of variance accounted for in a relationship
- coefficient of multiple correlation** An indicator of the combined relationship of several independent variables with the dependent variable
- Cohen's kappa** A type of interrater reliability for categorical data
- collective case** A type of case study design in which more than a single example or setting is used
- comparative** See comparative research

- comparative research** A type of nonexperimental quantitative research that examines differences between groups
- complete observer** An observer who remains completely detached from the group or process of interest
- comprehensive sampling** The type of sampling in qualitative research in which every participant, group, setting, event, or other information is examined
- concept analysis** A study that clarifies the meaning of a concept by describing its generic meaning, different meanings, and appropriate use
- concurrent triangulation design** A type of mixed method design in which quantitative and qualitative methods are used simultaneously
- confidence interval** A range that describes probable population values
- confidentiality** Only the researcher has access to the data and participants' names and subjects know beforehand who will see the data
- construct** A complex abstract concept that is not directly observable, such as anxiety, intelligence, and self-concept
- construct irrelevant variance** The extent to which an assessment measures facets that are not related to its purpose
- construct underrepresentation** The extent to which an assessment fails to incorporate important facets that are related to its purpose
- construct validity** A type of experimental validity that refers to the extent to which a study represents the underlying construct
- context sensitivity** In qualitative research, integrating aspects of the context in conducting the study and interpreting the results
- continuous observation** An observational data-gathering technique in which the observer records all important behaviors
- continuous variable** A variable in which the property or attribute of an object, subject, or entity is measured numerically and can assume an infinite number of values within a range; also called a *measured variable*
- control group interrupted time-series design** A quasi-experimental time-series study that compares the intervention group to a control group
- control or comparison group** Group of subjects in an experiment compared to the intervention group
- convenience sampling** A nonprobability method of selecting subjects who are accessible or available
- correlation coefficient** A calculation that represents the size and direction of the degree of relationship between two variables
- correlational research** Research in which information on at least two variables is collected for each subject in order to investigate the relationship between the variables
- cost-effective analysis** An evaluation that compares the outcomes of similar programs and practices in relation to their costs when the programs have the same objectives and measures
- credibility** The extent to which the results of a study approximate reality and are thus judged to be trustworthy and reasonable
- criterion variable** In a prediction study, the variable that is predicted
- criterion-referenced** Refers to instruments whose scores are interpreted by comparing them to set criteria or standards rather than to the performance of others
- critical studies** Qualitative research in which the researcher is committed to exposing social manipulation and changing oppressive social structures and in which he or she may have emancipatory goals
- Cronbach's alpha** A measure of internal consistency reliability for items with scaled responses
- cross-sectional** Refers to a research strategy in which several different groups of subjects are assessed at the same time
- crystallization** An analytical style in which the researcher combines segmenting, categorizing, and pattern seeking into an extensive period of intuition-rich immersion in the data
- culture** Shared norms and expectations for behavior among members of the same group
- data** The results obtained by research from which interpretations and conclusions are drawn
- deception** A procedure in which participants are not informed of the actual purpose of the study
- decision-oriented evaluation** An evaluation that supplies information for prespecified decisions, such as needs assessment, program planning, program implementation, and outcomes
- degrees of freedom** A mathematical concept that indicates the number of observations that are free to vary
- demand characteristics** A possible source of bias when any aspect of a study reveals its purpose and may influence subjects to respond differently because they know that purpose
- dependent variable** The measured variable that is the consequence of or depends on antecedent variables
- descriptive** Refers to research that describes an existing or past phenomenon in quantitative terms
- descriptive statistics** Statistical procedures used to describe something
- diffusion of intervention** A threat to internal validity in which the subjects are influenced by other conditions of the independent variable
- discrepant data** In qualitative studies, evidence that some data are not in agreement with other data
- documents** Records of past events, whether written or printed, such as letters, diaries, and journals, newspapers, and regulations
- double-barreled questions** Single questions that contain two or more ideas to which the subject must make one response
- duration recording** A type of observer recording procedure in which the duration of behavior is recorded
- ecological external validity** The extent to which the results of research can be generalized to other conditions and situations
- effect size** A statistical index of the practical or meaningful differences between groups
- emergent design** A research plan in which each step depends on the results of the field data obtained in the previous step
- enquiring** Action research in which new data are collected
- empirical** What is guided by evidence, data, and sources
- equivalence** A type of test reliability in which the scores from equivalent or parallel forms of the same instrument, obtained at about the same time, are correlated
- ERIC (Educational Resources Information Center)** A comprehensive database and index of education literature
- ethnography** A description and interpretation of a culture, social group, or system
- evaluation** A study that uses a formal design to collect and analyze data about a practice or anticipated practice and then determines the worth of that practice

- evaluation approach** A strategy used to focus evaluation activities and produce a useful report
- evidence based on contrasted groups** Validity evidence based on scores from groups expected to show differences
- evidence based on internal structure** Validity evidence that shows appropriate correlations among items
- evidence based on relations to other variables** Validity evidence that shows appropriate correlations with other measures
- evidence based on test content** Validity evidence in which scores represent an underlying meaning, interpretation, trait, or theory
- examining** Action research in which already collected data are analyzed
- exempt** Category or review used by Institutional Review Boards (IRB) to designate that proposed studies are not subject to procedures to obtain informed consent or continuing IRB review.
- expedited** Category of review used by Institutional Review Boards (IRB) to designate that proposed studies having no more than minimal risk to participants do not need full IRB review
- experiencing** Action research with a focus on observation to better understand the participants and the setting
- experimental design** Research in which the independent variable is manipulated to investigate a cause-and-effect relationship between it and the dependent variable
- experimenter effects** A threat to internal validity in which the researcher's differential treatment of the subjects affects results, also called *experimenter contamination*
- explanatory design** Mixed method design in which quantitative data gathered first, followed by qualitative data
- exploratory design** Mixed method design in which qualitative data gathered first, followed by quantitative data
- experimental group** See intervention group
- ex post facto design** See ex post facto research
- ex post facto research** Research that investigates events that have already occurred and implies a cause-and-effect relationship from the results
- external criticism** Analytical procedures carried out to determine the authenticity of the source—that is, whether the source is the original document, a forged document, or a variant of the original document
- external validity** The extent to which the results of a study can be generalized to other subjects, conditions, and situations
- extraneous variable** Uncontrolled variable that influences research results
- factorial ANOVA (analysis of variance)** An analysis of variance statistical procedure using two or more independent variables that permits testing each independent variable and the interaction among the variables
- factorial designs** Research in which there are at least two independent variables that are analyzed together
- facts** In analytical research, descriptions of who, what, when, and where an event occurred; obtained from decisive evidence
- fidelity of intervention** The extent to which an experimental intervention was completed as planned
- field log** Documents the nature, dates, and duration of work in the field
- focus group interview** A small group interview of selected individuals to assess a problem, concern, new product, program, or idea
- foreshadowed problems** Anticipated research problems that will be reformulated during data collection
- formative evaluation** Evaluation that is used to improve an ongoing practice or program
- frequency-count recording** A type of observer recording procedure in which the frequency of a behavior is recorded
- frequency distribution** A display of a set of scores that is organized by the number of times each score was obtained
- frequency polygon** A graphic representation of a frequency distribution formed by connecting in a line the highest frequency of each score
- gatekeeper** The individual(s) who provide(s) access to the field in qualitative and mixed method studies
- generalization** The extent to which the results of one study can be used as knowledge about other populations and situations; also, summaries of facts
- grounded theory** Qualitative procedures that are used to develop detailed concepts or conditional propositions for substantive theory
- high inference** A type of observation in which the observer records judgments about what has occurred
- histogram** A graphic illustration of a frequency distribution in which a bar is used to represent the frequency of each score
- historical analysis** The application of analytical methodology to the study of the past, as in biographies and studies of movements, institutions, and concepts
- historiography** A study of the procedures that different historians use in their research; also a study of the changing revisions and interpretations of the past
- history** A threat to internal validity in which incidents or events that occurred during the research affect results
- inadequate preoperational explication of the constructs** A threat to the construct validity of a study in which insufficient explanation is provided of the nature of the construct being measured or manipulated
- independent samples *t*-test** An inferential statistical procedure for determining the probability level of rejecting the null hypothesis using two samples of subjects that have no relation to each other
- independent variable** A variable that is antecedent to or that precedes the dependent variable; in experimental design, also called the *experimental* or *manipulated variable*
- in-depth interviews** Purposeful conversations that use a general interview guide with a few selected topics and probes (i.e., not a set of standardized questions); should last for at least an hour
- inductive reasoning** An analysis in which categories and patterns emerge from the data rather than being imposed on them prior to data collection
- inferential statistics** Procedures that are used to indicate the probability associated with saying something about a population based on data from a sample
- informed consent** Obtaining permission from individuals to participate in research before the research begins
- Institutional Review Board (IRB)** An organization that reviews research involving human subjects to ensure that ethical and legal practices are followed
- instrumental case** A type of case study design in which the focus is on a specific theme or issue
- instrumentation** A threat to internal validity in which changes in instruments and unreliability affect the results
- interaction** The unique effect that different levels of independent variables have on the dependent variable
- internal consistency** A type of test reliability in which the homogeneity of the items of an instrument is assessed after it has been administered once

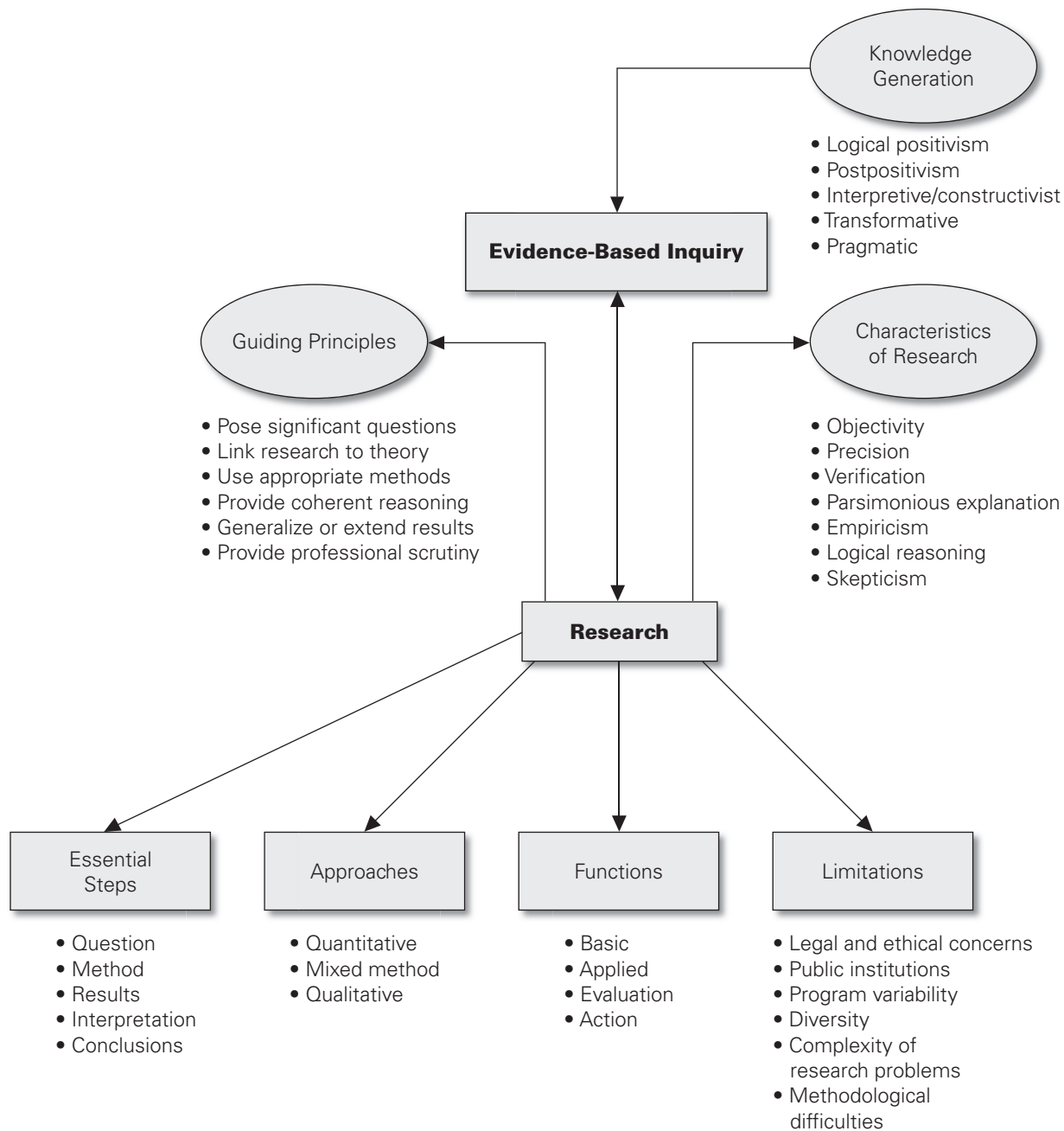
- internal criticism** The use of analytical procedures to determine the credibility of the statements in a source; the accuracy and trustworthiness of the facts
- internal validity** The degree to which extraneous and confounding variables are controlled
- interpretive/constructivist** Use of systematic procedures with multiple socially constructed realities
- interval** A type of measurement scale in which numbers are rank ordered with equal intervals between ranks
- interval recording** A type of observer recording procedure in which behavior that occurs during a given time interval is recorded
- intervention group** The group of participants that receives the intervention
- intervention, experimental, or treatment group** Group of subjects in an experiment who receive the intervention the researcher hypothesizes will change thinking or behavior
- intervention replications** A threat to internal validity in experiments that occurs when the number of intervention replications does not equal the number of subjects
- intrinsic case** A type of case study design in which the emphasis is on the case itself
- in vivo codes** In qualitative data analysis, the coding of participants' voices from the data
- Kuder-Richardson (KR)** A type of internal consistency reliability for items that are scored right or wrong
- kurtosis** A distribution that is either more peaked or more flat than a normal distribution
- leading question** A question that suggests a particular answer
- level of significance** A value that is selected to indicate the chance that it is wrong for the purpose of rejecting the null hypothesis; also called *level of probability* or *level of confidence*
- Likert scale** A type of scale in which the subject expresses a degree of agreement or disagreement with a statement
- loaded question** A question for which there is an obvious or desirable answer
- logical positivism** A rationalistic view of the world holding that humans can be studied like nature, with objective methods, with a single reality
- logistic regression** A type of regression analysis in which the dependent variable is dichotomous
- longitudinal** A research strategy in which quantitative data are collected on subjects over a period of time
- low inference** A type of observation in which the observer records the occurrences of specific behaviors
- magnitude of effect** A measure of the extent to which findings have practical significance
- mapping the field** Identifying the social, spatial, and temporal relationships among participants and the setting
- margin of error** Range of likely population values from the sample mean
- matrix sampling** A type of sampling in which parts of the test items are randomly assigned to each subject
- maturation** A threat to internal validity in quantitative research in which maturational changes in the subjects (e.g., growing older or becoming tired or hungry) affect the results
- maximum variation sampling** Type of qualitative sampling in which participants are selected to represent differences on characteristics of interest
- MAXMINCON** An acronym for maximizing systematic variance, minimizing error variance, and controlling extraneous variance
- mean** A measure of central tendency; the arithmetical average of the scores
- measurement scales** Properties that describe the relationships between numbers
- measures of central tendency** Summary indices of a set of scores that represent the typical score in a distribution
- measures of variability** Numerical indices that describe the degree of dispersion of scores from the mean
- median** A measure of central tendency; the point or score in a distribution that is the midpoint
- member checking** Participant review of notes and recordings for accuracy
- meta-analysis** A research procedure that uses statistical techniques to synthesize the results of prior independently conducted studies
- metasearch engines** Research tools that allow multiple Internet searches to be conducted at one time
- mixed method** Refers to a study that combines qualitative and quantitative techniques and/or data analysis within different phases of the research process
- mode** A measure of central tendency; the most frequently occurring score
- mono-method bias** A threat to construct validity due to the use of a single exemplar or measure
- mono-operation bias** A threat to construct validity due to the use of a single method in implementing an intervention or measuring the dependent variable
- multilevel mixed method sampling** Selection in which individuals are nested in larger groups
- multiple-baseline designs** A type of single-subject design that uses several subjects, types of behavior, or situations simultaneously
- multiple regression** A statistical procedure for using several variables to predict an outcome
- multisite evaluation** A type of qualitative research designed to report the practices at each site and to make generalizations across sites
- multistage cluster sampling** The use of several stages of clustering in selecting a sample
- multivariate** Refers to a family of statistics that are used when there is more than one independent variable and/or more than one dependent variable
- negative case** In qualitative research, a case that is found to contradict other cases
- negatively skewed** A distribution of scores that has a disproportionately large number of high scores
- nominal** A type of measurement scale in which objects or people are named, classified, or numbered
- nonequivalent groups posttest-only design** A pre-experimental design in which one or more groups of subjects (who have not been randomly assigned) receives an intervention and a posttest and another group of subjects receives only a posttest
- nonequivalent groups pretest–posttest control or comparison group designs** Quasi-experimental designs in which groups that have not been randomly assigned to interventions are compared using both a pretest and a posttest
- nonexperimental** Research that contains no direct manipulation of variables, such as descriptive and correlational research
- nonparametric** Types of statistical procedures used when the assumptions necessary to use parametric procedures have been violated
- nonprobability sampling** A sampling procedure in which the probability of selecting elements from the population is not known

- nonproportional sampling** Stratified sampling in which the number of subjects selected from each stratum is not based on the percentage of the population represented by that stratum
- norm-referenced** Refers to an interpretation of test results in which a score or group of scores is compared with the typical performance of a given (i.e., norm) group
- normal distribution** A symmetrical, bell-shaped distribution of scores that have the same mean, median, and mode
- null hypothesis** A formal statistical statement of no relationship between two or more variables
- objectives-oriented evaluation** An evaluation that determines the degree to which the objectives of a practice have been attained by a target group
- objectivity** Refers to data collection and analysis procedures from which only one meaning or interpretation can be made
- odds ratio** The nature of the results from a logistic regression that indicates the probability of some outcome
- open form** A type of questionnaire item in which the subject writes in a response to a question
- operational definition** A definition of a variable that is produced by specifying the activities or operations necessary to measure, categorize, or manipulate it
- oral history** A form of historical research in which individuals' spoken words and testimonies about the past are recorded
- oral testimonies** The records or interview transcripts of witnesses or participants to a past event that is being studied
- ordinal** Refers to a type of measurement scale in which the objects or persons are rank ordered from lowest to highest
- outlier** A data point that is extremely high or low and thus is very different from the other data collected
- paired *t*-test** An inferential statistical procedure for determining the probability level of rejecting the null hypothesis with two samples of subjects that are matched or related; also called *correlated samples* or *dependent samples *t*-test*
- parametric** Refers to types of statistical procedures that assume normality in population distributions, homogeneity of variance, and interval or ratio scale data
- participant observer** One who uses interactive data collection strategies such as limited participation, field observation, interviewing, and artifact collection
- participant-oriented evaluation** A holistic approach to evaluation that uses multiple methods to uncover the divergent values of a practice from the various participants' perspectives
- participants' perspectives** The language of participant descriptions
- path analysis** A statistical procedure that uses correlations among a set of variables that are logically ordered to reflect causal relationships
- pattern** A relationship among categories in qualitative analyses
- percentile rank** The point in a distribution at or below which a given percentage of scores is found
- phenomenological** Research that describes the meanings or essence of a lived experience
- plagiarism** Not giving credit to an original source of an idea or writing
- planned comparisons** Predetermined statistical tests of selected pairs of means
- plausible rival hypotheses** Possible explanations (i.e., other than the effect of the independent variable) for cause-and-effect relationships
- policy analysis** Research to investigate and formulate policies and programs
- population** A group of individuals or events from which a sample is drawn and to which results can be generalized
- population external validity** The extent to which the results of a research study can be generalized to other people
- portfolio** A form of alternative assessment in which the materials demonstrating student performance are purposefully collected, organized, and evaluated
- positionality** A researcher's display of position or standpoint by describing his or her own social, cultural, historical, racial, and sexual location in the study
- positively skewed** A distribution of scores that has a disproportionately large number of low scores
- post hoc comparison** Statistical tests used with pairs of means that are usually conducted after statistical test of all means together; also called *multiple comparisons*
- postpositivism** Allows for contextual limitations and determinants for logical positivism
- pragmatic** A research paradigm that includes common sense and practical thinking along with quantitative and qualitative methods
- prediction study** Research in which behaviors or skills are predicted by one or several variables
- predictor variable** The antecedent variable in a prediction study
- pre-experimental designs** Experimental designs that generally have very weak internal validity due to no pretest or no control group
- pretesting** A threat to internal validity in which taking a pretest affects the results
- primary source** In analytical research, a document or the testimony of an eyewitness to an event; in reviewing literature, studies with original data
- probability** A statement of the degree of confidence about predicting some outcome
- probability sampling** A type of sampling in which subjects are drawn from a population in known probabilities
- probing** Questions designed to lead to more detail in interviews
- propensity score matching** A statistical procedure in which participants are matched on a number of variables
- proportional sampling** A type of stratified sampling in which the number of participants selected from each stratum is based on the percentage of participants in the population in that stratum
- psycINFO** Database of psychological research and other literature sponsored by the American Psychological Association
- purposeful sampling** A type of sampling that allows choosing small groups or individuals who are likely to be knowledgeable and informative about the phenomenon of interest; selecting cases without needing or desiring to generalize to all such cases
- purposive random sampling** In mixed method studies, a small, targeted random sample
- qualitative** A type of research that refers to an in-depth study using face-to-face or observation techniques to collect data from people in their natural settings
- quantitative** A research paradigm in which objective data are gathered and analyzed numerically
- quasi-experimental designs** Research designs in which there is no random assignment of subjects
- questionnaire** A written set of questions or statements that is used to assess attitudes, opinions, beliefs, and biographical information

- quota sampling** A nonprobability method of sampling in which subjects are selected in proportion to the characteristics they represent in the general population
- random assignment** A procedure used to assign subjects to different groups so that every subject has an equal chance of being assigned to each group
- random sampling** A procedure for selecting subjects from a population in such a way that every member of the population has an equal chance of being selected
- randomized posttest-only control and comparison group design** A true experimental design in which one or more randomly assigned groups of subjects receives an intervention and a posttest and one randomly assigned group of subjects receives only a posttest
- randomized groups pretest-posttest comparison group design** Experiment in which there is random assignment, a pretest, a posttest, and at least two interventions are compared
- randomized groups pretest-posttest control group design** Experiment in which there is random assignment, a pretest, a posttest, and a control group
- range** A measure of variability; the difference between the highest and lowest scores in a distribution
- rank order** Listing of results from highest to lowest
- ratio** A type of measurement scale in which the numbers are expressed meaningfully as ratios
- realist ethnography** A detailed description of a cultural group
- recursive** In qualitative studies, findings that occur repeatedly
- reflexivity** Refers to the researcher's rigorous self-scrutiny throughout the entire qualitative research process
- reflex journal** See reflex records
- reflex records** Records the researcher makes immediately after leaving the field; contains summaries of observations, addresses quality of data, suggests next steps, and provides for self-monitoring
- regression coefficient** A factor used in multiple regression to weight the contribution of each variable in the equation
- relics** In historical research, objects that provide information about the past, such as textbooks, equipment, and examinations
- replication** A study that duplicates the findings of a prior study using different participants, settings or techniques
- research** A systematic process of collecting and logically analyzing data for a specific purpose
- research design** The plan that describes the conditions and procedures for collecting and analyzing data
- research methods** The procedures used to collect and analyze data
- research problem** A formal statement of the question or hypothesis that will be investigated through empirical research
- research role** The relationships acquired by and ascribed to the researcher during interactive data collection; should be appropriate for the purpose of the study
- response set** Tendency to respond the same way to multiple questions
- responsive evaluation** An evaluation designed to supply information about the issues and concerns of the audiences; uses an emerging design to provide an understanding of the program
- restriction in range** Small variation among scores
- reversal, removal, or withdrawn design** A type of single-subject design in which an intervention is changed or discontinued
- sample** The group of subjects from whom data are collected; often representative of a specific population
- sampling by case type** Sampling in qualitative research that depends on the type of participant that is selected, such as typical case or unique case
- sampling distribution** The frequency distribution of possible samples from a given population
- scale** Questionnaire items for which the responses consist of gradations, levels, or values that describe various degrees of something
- scatterplot** A graphic representation of the intersections of subjects' scores on two variables
- search engines** Services that allow for cataloging and retrieving information from the Internet
- secondary data** Data that were collected previously and are available in a database for further use
- secondary data analysis** Statistical analysis that uses secondary data
- secondary sources** In historical research, documents and testimonies of individuals who did not actually observe or participate in the event being studied; in literature reviews a summary of primary sources
- segment** A part of a dataset that is comprehensible by itself and contains one idea, episode, or piece of information relevant to the study
- selection** A threat to internal validity in which differences between groups of subjects affect the results
- semantic differential** A type of scale in which subjects respond by choosing between adjective pairs in relation to a concept or object
- semistructured questions** Fairly specific interview questions that allow for individual, open-ended responses
- sequential explanatory design** A mixed method design in which a quantitative phase is followed by a qualitative phase
- sequential exploratory design** A mixed method design in which a qualitative phase is followed by a quantitative phase
- sensitivity** The ability of interventions and measures to show relationships
- simple random sampling** A sampling method in which every member of the population has the same chance of being selected
- single-group interrupted time-series design** A quasi-experimental design in which multiple observations of the dependent variable are made before and after the intervention
- single-group posttest-only design** A pre-experimental design in which a single group of subjects receives an intervention and a posttest
- single-group pretest-posttest design** A pre-experimental design in which a single group of subjects receives a pretest, an intervention, and then a posttest
- single-sample t-test** An inferential statistical test of the difference between the mean of a set of scores and a set value
- single-subject designs** Research done with individual subjects in order to study the changes in behavior that are associated with the intervention or removal of the intervention
- site selection** The specification of site criteria implied in the foregrounded problems; used to obtain a suitable and feasible research site
- skewed** See positively skewed and negatively skewed
- snowball sampling (network sampling)** A qualitative strategy in which each successive participant or group is named by a preceding group or individual

- social desirability** The tendency of subjects to respond to items in ways that will seem desirable to others
- Social Science Citation Index (SSCI)** Provides access to the bibliographic and citation information to find research data, trends, journals, and researchers
- sources of variability** Systematic, error, and extraneous influences related to research design
- split-half reliability** A type of internal consistency reliability in which equal halves of a test are correlated
- spurious correlation** A correlation that overrepresents or underrepresents the true relationship
- stability** A type of test reliability that correlates scores from the same instrument given on two occasions
- stakeholder** A person, organization, or group that is interested in or impacted by the evaluation
- standard deviation** A measure of variability; a numerical index that indicates the average dispersion or spread of scores around the mean
- standard error** The standard deviation of a sampling distribution
- standard scores** Numbers that have been converted from raw distributions with constant means and standard deviations
- standardized tests** Tests that are administered and scored according to highly structured, prescribed directions
- standards-based** A type of test in which performance is compared to set standards of proficiency
- statistical conclusion validity** The extent to which statistics provide accurate information about the relationship being studied
- statistical hypothesis** A hypothesis that is stated in terms of statistical results
- statistical power** The ability of a statistical analysis to detect relationships with a given variance and a given number of participants
- statistical regression** The tendency for extreme scores to move closer to the mean score on a second testing
- statistically significant** Refers to evaluating the results of inferential statistics and indicating that the differences noted are not likely due to chance
- statistics** Procedures for organizing and analyzing quantitative data
- stem-and-leaf display** A method of showing a frequency distribution
- stratified purposeful sampling** In mixed method studies, targeted selection from each stratum
- stratified random sampling** A form of random sampling in which a population is first divided into subgroups (i.e., strata) and then subjects are selected from each subgroup
- structural equation modeling** Statistical method for using correlational relationships among latent variables to explain causal conclusions
- structured questions** Types of interview questions that provide a predetermined set of responses from which the participant is to choose; also called *limited-response questions*
- subject directories** Lists of cataloged Internet resources
- subject effects** Changes in subject behavior that result from being in a study
- summative evaluation** An evaluation designed to determine the merit, the worth, or both of a developed practice and to make recommendations regarding its adoption and widespread use
- survey research** The use of a questionnaire or interview to assess the current opinions, beliefs, and attitudes of members of a known population
- systematic sampling** A form of sampling in which subjects are selected from a continuous list by choosing every *n*th subject
- t-test** An inferential statistical procedure for determining the probability level of rejecting the null hypothesis that two means are the same
- target group** The group whose behavior is expected to change as a result of a given practice
- teacher-researcher** Teacher engaged in action research
- test reliability** The extent to which scores from an instrument are consistent
- test validity** The extent to which inferences based on instrument scores are reasonable
- theory** A prediction and explanation of natural phenomena
- thesaurus** A publication that lists and cross-references the key terms used in an index for a reference service (database), such as ERIC or *Psychological Abstracts*
- time sampling** A type of observer recording procedure in which behaviors are observed for specific time periods
- time-series design** Quasi-experimental design in which one group of subjects is measured repeatedly before and after an intervention
- transformative** A research paradigm in which social, political, cultural, racial, and ethnic factors contribute to the design and interpretation of results
- treatment group** See intervention group
- triangulation** Qualitative cross-validation among multiple data sources, data collection strategies, time periods, and theoretical schemes
- triangulation design** A mixed method design in which quantitative and qualitative data are collected at the same time
- true experimental** A type of experimental research design that uses random assignment of subjects to different groups
- Type I error** The error that results from rejecting the null hypothesis when it is in fact true
- Type II error** The error that results from failing to reject the null hypothesis when it is in fact false
- typicality** The degree to which a phenomenon may be compared or contrasted with other phenomena along relevant dimensions
- unit of analysis** Smallest element used as the number of replications employed in statistical analyses.
- unit of study** Independent element that corresponds to the number of participants
- univariate** Refers to a statistical analysis in which there is a single dependent variable
- unobtrusive measures** Methods of collecting information in which the subject is unaware of being a participant in the research; also called *nonreactive measures*
- unstructured questions** Interview questions that are broad and allow for open-ended responses
- variability** See measures of variability
- variable** An event, category, behavior, or attribute that expresses a construct and has different values, depending on how it is used in a study
- variance** Generically, the degree of spread or dispersion of scores; mathematically, the square of the standard deviation
- verification** Confirming or modifying the results of a research study in subsequent research
- visual representation** An organized assembly of information (e.g., figures, matrices, integrative diagrams, and flow charts) that assists in qualitative data analysis
- z-score** A type of standard score that has a mean of 0 and a standard deviation of 1

Introduction to Evidence-Based Inquiry



KEY TERMS

logical positivism	evidence-based inquiry	mixed method
postpositivism	generalization	basic research
interpretive/constructivist	research	theory
quantitative	research methods	applied research
qualitative	objectivity	evaluation research
transformative	verification	action research
pragmatic	empirical	
replication	data	

WHAT YOU WILL LEARN

Study this chapter and you will:

- Understand that the importance of educational research has been heightened by federal efforts to make educational research more like scientific research.
- Know how enhanced technology has made research much more accessible.
- Explain how research is fundamentally based on human judgments.
- Explain how evidence-based knowledge can make significant improvements in the practice of education.
- Distinguish between various epistemologies that influence research methods.
- Understand that educational research is systematic inquiry guided by established principles and characteristics.
- Distinguish between research methods that are quantitative, qualitative, or a combination of the two.
- Know the five steps required to conduct research.
- Distinguish between quantitative, qualitative, and mixed methods approaches to research.
- Know the differences between basic, applied, evaluation, and action studies.
- Know the constraints and limitations of educational research.

EDUCATIONAL RESEARCH IN THE TWENTY-FIRST CENTURY

The times we live in are truly amazing in terms of the possibilities for educational research! Powerful tools have been afforded us through the variety of technology and research methods that have been refined throughout the past half century. These tools and methods allow us to address challenging questions and to have greater confidence that our results will be valid and useful. More important, there is a renewed interest at all levels of education for decisions to be data driven and based on hard evidence. This has resulted in a greater need for all educators to understand, conduct, and use research findings.

While educational research has been conducted for decades, two developments at the federal level have significantly influenced the nature of what is researched and how studies are conducted. In 2002, based on *No Child Left Behind*, President George W. Bush signed the Education Scientific Reform Act. This legislation led to the development of the Institute for Education Sciences (IES). This important office has the following mission:

The mission of the Institute is to provide national leadership in expanding fundamental knowledge and understanding of education from early childhood through post-secondary study, in order to provide parents, educators, students, researchers, policymakers, and the general public with reliable information about the condition of and progress of education . . . educational practices that support learning improve academic achievement and access . . . [and] the effectiveness of Federal and other education programs.

Source: From H. R. 3801 (2002). An act to provide for improvement of federal education research, statistics, evaluation, information, and dissemination and for other purposes. (p. 5)

At about the same time, the National Research Council (NRC) published *Scientific Research in Education* (Shavelson & Towne, 2002). This book lays out a series of principles to guide research (we'll consider these shortly), as well as serving as a set of principles for judging the adequacy of

empirical studies. Together, these two developments have had a strong, ubiquitous influence on educational research. They have had a direct impact on what is studied, how studies are conducted, and how results are reported and disseminated. Most importantly, there is a new attitude about the importance of scientific or evidence-based research as the basis for establishing knowledge about what educational practices have the greatest impact. Consistent with No Child Left Behind, there is an emphasis on outcomes—in this case, scientifically based results.

The other major development affecting educational research is technology. Technology has changed how research is conducted, such as the use of Web-based surveys and statistical software, making it very easy to analyze data. Online journals, websites, and search engines also provide amazingly fast access to studies and other literature.

Although these developments may suggest a more objective approach to educational research, we need to be clear about the role of human judgment. Many would argue that, in the end, all of our research and the knowledge generated from that research are fundamentally based on subjective premises. For example, isn't it the researcher who decides how to frame survey questions? Isn't the choice of statistical analysis up to the researcher? What about the interpretation of results? Is *cooperative learning* defined the same way for different studies? Our perspective is that human judgment is indeed critical to research, but we also believe that there are principles of evidence-based thinking that make such judgments more accurate. After all, at the end of the day, we want information that will enhance student motivation and achievement and improve attitudes. Research framed as evidence-based inquiry helps us generate more credible information. We'll expand on this last point in the next section.

Why Educational Research Is Important

Why has educational research become a valuable source of information? We suggest six reasons for the importance of evidence-based inquiry.

First, *educators are constantly trying to understand educational processes and must make professional decisions*. These professional decisions have immediate and long-range effects on others: students, teachers, parents, and, ultimately, our communities and nation. How do educators acquire an understanding to make decisions? Most of us tend to rely on several sources, including personal experience, expert opinion, tradition, intuition, common sense, and beliefs about what is right or wrong. Each of these sources is legitimate in some situations, yet in other situations, each source may be inadequate as the only basis for making decisions.

Second, *noneducational policy groups, such as state and federal legislatures and courts, have increasingly mandated changes in education*. How do policy groups acquire their views of education and obtain their information about schools and instruction? Most policy-makers prefer to have research-based information relevant to the specific policy issue. Many state legislatures require state education departments to conduct studies on state educational policies. Both federal and state departments of education also commission funded studies. Researchers are increasingly being asked to work on complex problems in highly politicized environments.

Third, *concerned public, professional, and private groups and foundations have increased their research activities*. Professional educational associations, teacher labor unions, Parent-Teacher Associations, and foundations such as the National Science Foundation have conducted or commissioned studies on topics of special concern to the organization.

Fourth, *reviews of prior research have interpreted accumulated empirical evidence*. For example, research reviews have addressed such topics as thinking aloud and reading comprehension; hypermedia and learner comprehension, control, and style; why parents become involved in their children's education; parameters of affirmative action in education; teacher efficacy; the effects of single-sex and co-educational schooling on social, emotional, and academic development; and teacher occupational stress, burnout, and health. Other research reviews identify areas of needed research.

Fifth, *educational research is readily available*. Research about educational practices is found in professional and research journals, funding agencies' published reports, books, library databases, newspapers, television, and the Internet. Although the quality of the research may vary with the specific source, educational research is very accessible.

Sixth, *many educators who are not full-time researchers conduct studies to guide their decisions and to enhance classroom, school, and system accountability*. Teachers can conduct action research

that is relevant for their needs and for the issues about which they feel passionately, such as second-language students, students with disabilities, and approaches to teaching school subjects. Educators often collaborate to conduct research and to form partnerships in projects. Seemingly insignificant findings can add to the current body of evidence in the search for answers to important educational questions. Furthermore, all educators must be able to demonstrate effectiveness in an age of accountability. Educators also need to interpret results accurately and to be responsible in their use of research findings. Evidence-based inquiry provides valid information and knowledge about education that can be used to make informed decisions.

Because research systematically describes or measures phenomena, it is often a better source of knowledge than one’s own experiences, beliefs, traditions, or intuition. Some studies provide general information about common educational practices and policies; this type of research influences the way one thinks about education. Other studies provide detailed information about specific practices at particular sites, such as a school, a classroom, or an office; this type of research can be used immediately to improve or justify a specific practice.

Using Evidence-Based Knowledge to Improve Educational Practices

The impact of educational research on schools and policy-makers seeking to improve educational practices may be seen as a process. Figure 1.1 shows the five phases of the process of developing evidence-based knowledge to improve educational practices: (1) identification of research problems, (2) empirical studies, (3) replications, (4) research synthesis and review, and (5) practitioner adoption and evaluation. The identification of research problems (Phase 1) begins with determining valued outcomes. Practical fields, like education, are concerned with valued outcomes such as learning. Research questions and problems come from the following sources: common observation, practical wisdom, policy controversies, prior research, and new methods applied in the study of education. Researchers conduct evidence-based studies (Phase 2), and then they attempt research replication (Phase 3) with different subjects and in a variety of settings and circumstances. In research synthesis and review (Phase 4), comparable studies are systematically evaluated and statistically or narratively summarized. Such an analysis helps to organize and make sense of the overall findings of prior research. Thus, the preponderance of evidence from many careful fully conducted studies, rather than a few exact replications of the original research, builds an evidence-based body of knowledge in education. Practitioners and policy-makers can reasonably accept the implications of research findings that are consistent without harmful side effects. Continuing local evaluation (Phase 5) is the final phase in the process.

To illustrate the potential impact of evidence-based research on educational outcomes, here are some examples of practices that were found to be effective (National Center for Educational Evaluation and Regional Assistance, 2004):

- **One-on-one tutoring by qualified tutors for at-risk readers in grades 1–3** The average tutored student read more proficiently than approximately 75% of the untutored students in the control group.
- **Life-skills training for junior high students** Implementing a low-cost, replicable program reduced smoking by 20% and a serious level of substance abuse by 30 percent by students’ senior year compared to the control group.

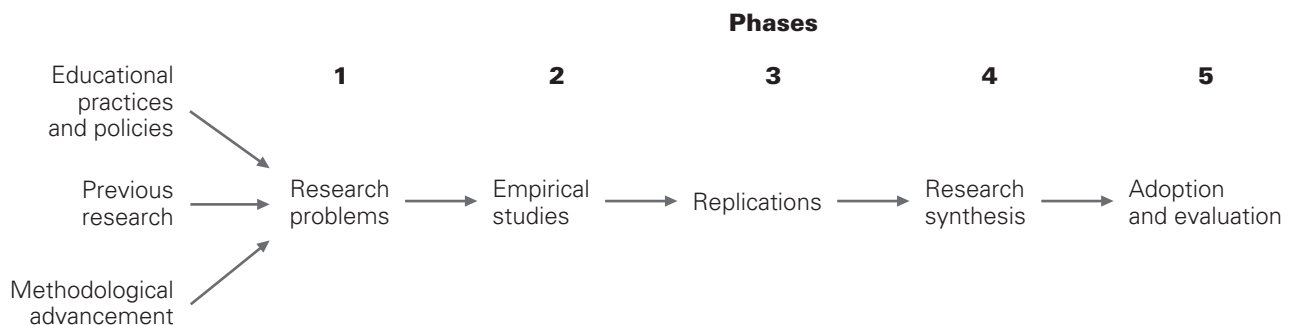


FIGURE 1.1 Development of Evidence-Based Knowledge to Improve Educational Practice

TABLE 1.1 Topics Researched by the NCEE

School choice	Magnet school assistance programs
Technology	Preschool reading
School-based drug use and violence prevention	Reading instruction in the primary grades
Mentoring	Teacher quality and preparation
Career and technical education	Professional development
After-school programs	English language learning
Charter schools	Remedial reading
Comprehensive technical assistance centers	

- **Reducing class size in grades K–3** On average, students in small classes scored higher on the Stanford Achievement Test in reading/math than 60 percent of the students in regular-sized classes.
- **Instruction for early readers in phonemic awareness and phonics** The average student in these interventions read more proficiently than approximately 70 percent of the students in the control group.
- **Enhancing students' literacy and early reading skills** Early Reading First students' average score of 102.69 was slightly above the national average for the print and letter knowledge subtest.

Clearly, the federal emphasis is on supporting a new generation of rigorous research.

The National Center for Educational Evaluation and Regional Assistance (NCEE) was established as part of the IES's goal to focus on evidence-based program effectiveness and impact questions. As illustrated in Table 1.1, the NCEE has initiated program effectiveness studies in areas that are directly related to student achievement. All of these nationally funded evaluation studies take either two or three years to address the bottom-line question of causality: Was the program or specific intervention effective? The NCEE supports only those studies that can provide credible scientific evidence. The results of these studies are widely disseminated in scholarly and professional journals and are available on the NCEE website. As evidenced by the number of different areas studied, many topics, issues, and programs are now viewed through the evidence-based lens of educational research.

RESEARCH AS SCIENTIFIC, EVIDENCE-BASED INQUIRY

Conducting research is a relatively new activity in the history of education. In the centuries before reading and writing were common, individuals developed knowledge of the world around them primarily by three means. The first was through personal experiences and observation of others' experiences. Collective wisdom was conveyed as a series of detailed stories of people and events. Stories provided an understanding, a repertoire of wisdom from which one could extrapolate or apply known experience to an unknown area and thus form reasonable expectations.

A second method of knowledge generation could be identified as logical positivism. This approach emphasized that there is a single reality within known probability, objectivity, empiricism, and numbers. **Logical positivism** was the foundation for the scientific method. The idea was that the study of humans could be done the same way as the study of nature, with an accepted set of rules for conducting and reporting the results. This rationalistic view of knowledge is illustrated by the work of Francis Bacon, John Locke, and August Comte. Prior to World War II, logical positivism gave way to **postpositivism**, which allows for limitations, contextual factors, and use of multiple theories within which research findings are interpreted.

Logical positivism: a rationalistic view of knowledge with single realities

Postpositivism: allows for limitations to rationalism and contextual factors