3

What Causes Measurement Error?

Overview

What causes measurement error? An understanding of different sources that can cause errors is important in trying to minimize errors to begin with. Whereas Chapter 2 discussed several types of errors, Chapter 3 discusses several sources that result in these errors. This chapter cross-lists *sources* of errors mentioned in the literature and likely *outcomes* in terms of different types of measurement error. Although this listing is the result of a detailed examination of error sources discussed in the social sciences literature, it is intended to be illustrative and not exhaustive. An understanding of the relationship between sources of error and the specific nature of measurement error they are likely to cause can be used to minimize error before the fact by designing appropriate items and measures.

Sources of Measurement Error

A variety of sources can cause measurement error, including response styles, specifically acquiescence, disacquiescence, extreme response, response range, midpoint responding, and noncontingent responding (Baumgartner & Steenkamp, 2001; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Acquiescence bias occurs when individuals differ in their tendency to agree with item statements. Such bias adds individual variation over and above variation in the construct being measured (Table 3.1). Agreement bias is a tendency to agree with statements, irrespective of the content of the item. Also referred to as acquiescence response style (Martin, 1964), it can be caused by several factors (Baumgartner & Steenkamp, 2001): stimulus-seeking extroverts (Couch & Keniston, 1960; Messick, 1991); lower status or cognitive ability of respondents (Knowles & Nathan, 1997; Messick, 1991; Schuman & Presser, 1981); ambiguous, vague, or neutral items (Messick, 1967; Paulhus, 1991; Ray, 1983); or distraction, time pressure, or other such factors (McGee, 1967). If higher levels of agreement occur because of pervasive factors such as wording, then this pattern leads to additive systematic error. If individuals vary consistently in their tendency to agree over and above the construct being measured, then this pattern leads to within-measure correlational systematic error. Disacquiescence response style, also referred to as disagreement bias or nay-saying, is the opposite of acquiescence response style and could be caused by stimulus-avoiding introverts (Couch & Keniston, 1960). Net acquiescence response style (Baumgartner & Steenkamp, 2001) is the sum of these two response styles and is also referred to as direction bias. In studies of response style effects (Baumgartner & Steenkamp, 2001; Martin, 1964), such response styles are assessed by the degree of agreement (or disagreement) with heterogeneous items from multiple scales without much in common, or from the extent of agreement with positively and negatively worded items from one scale before reverse scoring.

Location bias occurs when individuals differ in the manner in which they use response scale categories (e.g., a tendency to scale upward or use extremes). Leniency is the tendency of a respondent to rate too high or too low. Severity (or stringency) is the opposite of leniency. Midpoint responding, a tendency to use the middle scale point irrespective of content (Baumgartner & Steenkamp, 2001), may be caused by evasiveness, indecision, or indifference (Messick, 1968; Schuman & Presser, 1981). This response style has been measured by the proportion of use of midpoints (Chen, Lee, & Stevenson, 1995; Stening & Everett, 1984). Midpoint responding leads to use of middle alternatives and is likely to cause within-measure correlational systematic error. In other words, although not affecting means, this type of error can lead to consistent differences over and above the construct in question. If any such error affects a small proportion of individuals, it may be identified as idiosyncratic random error. However, if a factor such as the use of extreme wording in the end anchors has a more pervasive effect, then the result is within-measure correlational systematic error.

(Text continues on page 141)

Table 3.1 Incomplete Taxonomy of Error Sources and Errors

Sources of Error	*Category of Measurement Error*						*Description*
	Random Error			*Systematic Error*			
	Idiosyncratic	*Generic* *Within Administration*	*Generic* *Across Administration*	*Additive*	*Within Measure Correlational*	*Across* Measure*	
Idiosyncratic individual-related							
Language difficulties	X						
Transient personal factors							Churchill (1979)
Mood	X						
Fatigue	X						
Memory/attention vacillations	X						Bagozzi (1984); Ghiselli (1964)
Mechanical/motor vacillations	X						Check in wrong box (Bagozzi, 1984; Churchill, 1979)
Noncontingent responding	X						Marsh (1987); Watkins and Cheung (1995)
Other idiosyncratic responses	X						All sources of error below that affect few respondents
Generic individual-related							
Individual differences in social desirability					X	X	Crowne and Marlowe (1964)
Charitability bias					X	X	Couch and Keniston (1960)
Faking good/faking bad					X	X	Meehl and Hathaway (1946)
(Dis)Acquiescence bias					X	X	Nunnally (1978); Lennox and Dennis (1994); Martin (1964); Ray (1983); Couch and Keniston (1960)

(Continued)

Table 3.1 (Continued)

Sources of Error	*Category of Measurement Error*						*Description*
	Random Error			*Systematic Error*			
	Idiosyncratic	*Generic*		*Additive*	*Within Measure Correlational*	*Across* Measure*	
		Within Administration	*Across Administration*				
Rater dispersion bias					X	X	Use of more extreme scores leading to higher standard deviation (Braucht, 1972; Greenleaf, 1992b; Wyer, 1969)
Extreme response style					X	X	
Standard deviation or response range					X	X	Tendency to use a wide or narrow range of responses (Greenleaf, 1992a; Hui & Triandis, 1985; Wyer, 1969)
Midpoint responding					X	X	Tendency to use the midpoint irrespective of content (Messick, 1968; Schuman & Presser, 1981)
Item content-related							
Lack of clarity of measures		X	X				Ambiguous wording, incomplete wording, poorly defined terms (Churchill, 1979; Fowler, 1993)
Estimation		X	X	X			
Leading questions				X			
Direction of wording effects					X	X	Cronbach (1946)
Common stem/similar wording					X	X	Lennox and Dennis (1994)
(Dis)Acquiescence bias				X	X	X	Nunnally (1978); Lennox and Dennis (1994); Martin (1964); Ray (1983); Couch and Keniston (1960)

Sources of Error	*Category of Measurement Error*							*Description*
	Random Error				*Systematic Error*			
	Idiosyncratic	*Within Administration*	*Generic*	*Across Administration*	*Additive*	*Within Measure Correlational*	*Across* Measure*	
Midpoint responding						X	X	Tendency to use the midpoint irrespective of content (Messick, 1968; Schuman & Presser, 1981)
Response format-related								
Rater location bias						X	X	
Rater dispersion bias						X	X	Use of more extreme scores (Braucht, 1972; Greenleaf, 1992b; Wyer, 1969)
Extreme response style						X	X	
Standard deviation or response range						X	X	Tendency to use a wide or narrow range of responses (Greenleaf, 1992a; Hui & Triandis, 1985; Wyer, 1969)
(Dis)Acquiescence or yea-/nay-saying					X	X	X	Tendency to agree or disagree with items (Greenleaf, 1992a; Hui & Triandis, 1985; Wyer, 1969)
Leniency/stringency					X	X	X	Consistently too positive/negative (Alliger & Williams, 1992)
Central tendency						X	X	Avoiding extreme scores (Guilford, 1954)
Unbalanced category labeling					X			
Confusing/ambiguous category labeling		X		X				
Number of response categories		X		X				Cox (1980)

(Continued)

Table 3.1 (Continued)

Sources of Error	*Category of Measurement Error*					*Description*
	Random Error		*Systematic Error*			
	Idiosyncratic *Within Administration*	*Generic* *Across Administration*	*Additive*	*Within Measure Correlational*	*Across* Measure*	
Administration related						
Learning/training				X	X	Ghiselli (1964)
Fatigue	X	X		X	X	Ghiselli (1964)
Distracting setting	X	X				
Interviewer biases			X	X	X	
Variations in administration		X				Interviewers who probe differently (Churchill, 1979)
Logical error in rating				X	X	Similar responses to items presupposed to be logically related (Bardo et al., 1982; Newcomb, 1931)
Proximity error				X	X	Similar responses to items close together (Stockford & Bissell, 1949)
Common instructions across measures					X	Lennox and Dennis (1994)
Halo				X	X	Similar responses to items related to general construct (Bardo et al., 1982; Thorndike, 1920)
Social desirability			X	X	X	Crowne and Marlowe (1964)
Experimenter expectancy			X	X	X	Cook and Campbell (1979)
Hypothesis guessing				X	X	Cook and Campbell (1979)
Differential augmentation/ attenuation					X	Similar methods increase (and dissimilar methods decrease) observed relationship between traits as the true relationship between traits increases (Bagozzi & Yi, 1991; Campbell & O'Connell, 1967)

*Across-measure additive and correlational systematic error are combined into a single category here.

Extreme response style refers to choosing extreme responses irrespective of content (Greenleaf, 1992b). It could be caused by several factors (Baumgartner & Steenkamp, 2001): an intolerance for ambiguity or dogmatism (Hamilton, 1968); anxiety (Hamilton, 1968); respondents lacking appropriate cognitive schemas (Shulman, 1973); or stimuli that are meaningful, important, or involving to respondents (O'Donovan, 1965). It has been measured by the extent of use of extreme categories (positive or negative) (Bachman & O'Malley, 1984). Response range is the tendency to use response categories in a narrow or wide range (Greenleaf, 1992a; Hui & Triandis, 1985; Wyer, 1969) and may be caused by factors similar to those that cause extreme response style (Baumgartner & Steenkamp, 2001). It has been measured by the standard deviation in an individual's responses across items (Greenleaf, 1992a; Hui & Triandis, 1985; Wyer, 1969). Noncontingent responding is the tendency to be careless, random, or nonpurposeful in responding (Baumgartner & Steenkamp, 2001; Marsh, 1987; Watkins & Cheung, 1995) and may occur because of lack of motivation. It has been measured by summing the absolute differences between pairs of items that are highly correlated and have similar means across respondents, and that are worded in the same direction (Bachman & O'Malley, 1984; Baumgartner & Steenkamp, 2001; Watkins & Cheung, 1995).

Halo effect also causes within-measure correlational systematic error and is a tendency to provide similar responses across items that are thought to be related. Again, such error, if restricted to a small proportion of individuals, is similar to idiosyncratic random error. Proximity error relates to similar responses to items in proximity and can also result in within-measure correlational systematic error. Nay-saying and yea-saying can also lead to error. If wording leads respondents to agree (disagree) to a greater degree, this could lead to additive systematic error. However, if yea- and nay-saying vary consistently across individuals (i.e., lead to individual differences in yea- and nay-saying), then the resulting error is within-measure correlational systematic error. In other words, individuals differ on yea-saying and nay-saying tendencies, and their responses reflect these tendencies over and above the construct being measured. Social desirability, a tendency to present oneself in a favorable light, can similarly lead to additive or correlational systematic error. Standard deviation error—a tendency to use a wide or narrow range of responses—can increase or reduce spread. If individuals vary consistently in standard deviation, this pattern leads to within-measure correlational systematic error (i.e., consistent differences across individuals over and above the construct being measured). If a constant, pervasive effect leads to greater or lesser spread, this pattern is also an example of within-measure correlational systematic error.

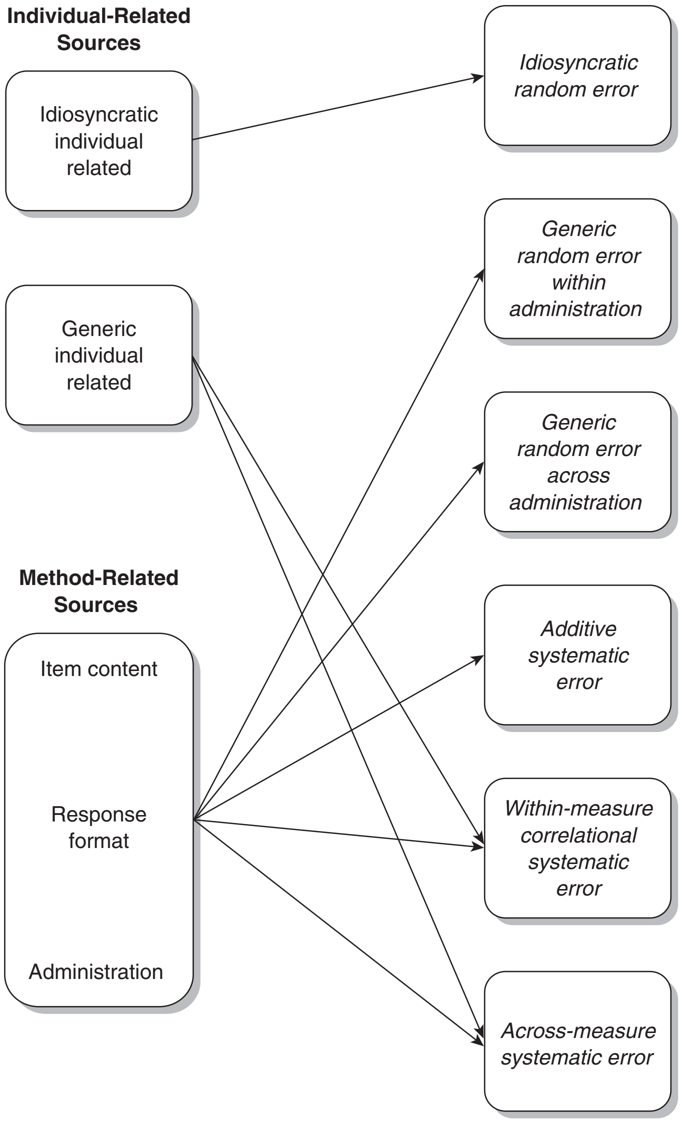

Figure 3.1 Summary of Error-Sources and Errors

In sum, within-measure correlational systematic error can be caused by consistent individual differences over and above the construct being measured. It results from underlying individual differences that lead to varying responses across individuals or pervasive wording or other factors that lead to constant dispersions across individuals, but reflect differences over and above the construct in question. It should be noted that additive systematic error is also constant across individuals but in one direction. Hence, it does not affect differences except when it reduces variance due to finite scale ends (i.e., partial additive systematic error).

Taxonomy of Error Sources

A taxonomy of sources of measurement error is developed here as a way to organize different sources (Table 3.1 and Figure 3.1). Past research lacks a taxonomy of a wide variety of error sources. The incomplete taxonomy lists several sources of errors mentioned in the literature and is intended to be illustrative and not exhaustive.

Past research has categorized sources of errors in several ways. Bardo et al. (1982) differentiate between respondent-related errors that are content-specific and form-related errors that are due to respondents' use of response formats. Method variance has been described as varying in level of abstraction from item wording and scale types to halo effects and social desirability (Bagozzi & Yi, 1991; Fiske, 1982). Ghiselli, Campbell, and Zedeck (1981) distinguish between situation-centered and person-centered sources of errors across administrations. However, past research is characterized by the lack of a taxonomy of a wide variety of error sources.

Several points are noteworthy about the taxonomy developed here. First, sources of error are differentiated from errors, an important and necessary distinction. Sources usually associated with random error can sometimes lead to systematic error and vice versa. Random and systematic measurement errors are outcomes of error sources. Second, no single category completely captures a source of error; hence, this taxonomy is necessarily an approximate one. At a broad level, all sources of errors arise out of the data collection event, which has been referred to as the union between a trait and a method (Campbell & Fiske, 1959).[1] A method is used here to refer to all the things that are done to collect the data, including the setting, administration procedures, and the measure itself. More specifically, sources of error may arise because of certain respondent characteristics, certain characteristics of the method, or interactions between these two categories.

The following categorization separates *individual-related* sources of errors (idiosyncratic and generic) from *method-related* sources of errors, with the latter being separated into item content, response format, and administration issues. Hence, there are certain characteristics (enduring and transient) that individuals bring into a data collection event, and there are certain characteristics of the data collection event. Each of these sets of characteristics can dominate in affecting responses. In addition, these sets of characteristics can interact to affect responses as well. The present classification focuses on categorization into single categories with the assumption that there is a likely dominant effect. Clearly, interactions between each of these components could lead to more complex categorization of error sources, and few such examples are provided. Despite the difficulty involved in clear categorization, such a taxonomy enables understanding of error sources and possible errors that can arise as consequences. Likely outcomes of error sources in terms of errors are identified here, fully recognizing that error sources could lead to errors other than the ones identified here.

Idiosyncratic Individual-Related Factors

Idiosyncratic individual-related factors are those that can affect a small proportion of individuals and include transient factors such as mood or language difficulties or distractions (Table 3.1). These sources of errors are usually idiosyncratic to individuals and are likely to lead to idiosyncratic random errors. They arise out of the state that the individual is in, such as an extreme mood, or are due to some idiosyncratic factors, such as mechanical variations. An individual's state could interact with some aspect of item wording or response format to lead to error. In other words, idiosyncratic individual factors could interact with aspects of the method to lead to error, but the outcome is idiosyncratic random error because it affects only a small proportion of individuals. Also, whether an error source is idiosyncratic or generic depends on how pervasive the error is. This distinction can be blurred as a function of the proportion of respondents affected. Moreover, if any of the other sources listed below affect a small proportion of individuals, they are indistinguishable from idiosyncratic individual-related factors.

Generic Individual-Related Factors

This category refers to individual differences along certain dimensions, such as social desirability, that are more pervasive than idiosyncratic factors (Table 3.1). Such factors, by their very nature, lead to correlational systematic error through the relationship between such individual differences and differences along a trait. They arise because of variations in the way individuals provide responses that are affected to different degrees for different individuals, over and above true differences in the construct being measured.

Hence, individual differences in impression management, charitability, or "faking good" may lead to correlational systematic error. It should be noted that whether individual-related factors are generic or idiosyncratic depends on how pervasive the factors are in influencing responses. Any generic factor could also be idiosyncratic in nature (e.g., faking good) if it affects only a small proportion of individuals, again highlighting the difficulty of separating out different sources of error.

Item Content-Related Factors

These factors relate to item wording effects, such as ambiguity and complexity, that may lead to pervasive errors (Table 3.1). A poorly defined word or term could lead to random error (e.g., "How much did you spend on recreational activities?"). Items to which respondents are unable to respond may lead to random error. Leading questions may lead to additive systematic error. Questions requiring estimation (e.g., "How many cans of Coke did you drink last year?") may lead to random error (because of guessing) or additive systematic error (because of underestimation or overestimation, say, because an inflated rate is computed based on purchase rather than usage by multiplying the cans purchased per week by the number of weeks in a year). Ambiguity could lead to use of the middle option and, hence, correlational systematic error that is coincidentally negatively related to the trait being measured. Again, it should be noted that if item-related factors affect a small proportion of individuals, they are indistinguishable from idiosyncratic individual-related factors.

Response Format-Related Factors

Response format-related factors have been included among what have been referred to in the literature as method factors (i.e., factors in the method employed that may cause responses). These response format-related factors include variations in the use of extremes or different parts of a scale (Table 3.1). They could lead to correlational systematic error either within or across measures, as well as additive systematic error and random error. Central tendency can cause correlational systematic error in that it is coincidentally negatively related to the trait being measured. Yea- and nay-saying tendencies may lead to additive systematic error. Unbalanced response categories (i.e., a set of response categories that does not have corresponding positive and negative levels, such as *excellent, very good, good, fair,* and *poor*) may lead respondents in one direction and cause additive systematic error. For instance, if most of the response categories in a scale are positive, they may cause responses to move toward the positive end of the scale. The close relationship between response format-related factors and generic individual-related factors is noteworthy. When enduring individual differences lead to responses being affected to different degrees, then the error source is categorized as generic individual-related and the outcome is correlational systematic error. When characteristics of the response format dominate and lead to dispersion or inflation/deflation (i.e., in one direction, or central tendency), then the error source is categorized under response format-related factors. Arguably, several of the error sources listed under this category could fit as interactions between generic individual-related factors and response format-related factors or under generic individual-related factors (e.g., a tendency toward yea-saying could be a generic individual difference or an interaction between individual- and response format-related factors, or it could be elicited by the response format). As discussed, sources such as acquiescence response style could be caused by individual differences (extroversion or cognitive ability), item content (ambiguity), or administration factors (time pressure). Potential interactions between individual differences and response format could lead to more complex forms of error. Again, it should be noted that if response format-related factors affect a small proportion of individuals, they are indistinguishable from idiosyncratic individual-related factors.

Administration-Related Factors

Administration-related factors include the setting, procedures (e.g., sequencing or administering items/measures contiguously), and interviewer/experimenter-related factors (e.g., leading on the part of the interviewer or experimenter). Interviewer biasing can lead to additive systematic error; it can also lead to correlational systematic error, for instance, because different respondents are differentially and consistently affected. Distracting settings and variations in administration are similar to ambiguous wording in terms of leading to generic random error. Similarly, logical error in rating (similar responses to items thought to be logically related) and halo effects can lead to correlational error within and across measures. Halo and proximity error sources are classified under administration-related factors because they arise out of items being administered together. Procedures can also elicit social desirability of an additive or correlational form, or even yea-saying tendencies and other sources listed under response format.

Interactive effects between traits and methods are also listed under administrative factors. Differential augmentation (Bagozzi & Yi, 1991; Campbell & O'Connell, 1967, 1982) occurs when "the higher the basic relationship between two traits, the more the relationship is increased when the same method is shared" (Campbell & O'Connell, 1982, p. 95). For instance, raters may have theories about how a pair of traits (say, value consciousness and price consciousness) is related. In such a situation, the stronger the true association between the traits, the more likely it is noticed and inflated (Bagozzi & Yi, 1991). Differential attenuation (Bagozzi & Yi, 1991; Campbell & O'Connell, 1967, 1982) occurs when "not sharing the same method dilutes or attenuates the true relationship, so that it appears to be less than it should be" (Campbell & O'Connell, 1982, p. 95). For instance, when collecting data on multiple occasions, the correlation between two related traits is attenuated for longer than for shorter intervals, whereas with two unrelated traits, no attenuation is possible (Bagozzi & Yi, 1991; Campbell & O'Connell, 1967, 1982).

The same sources could be categorized under item content-related factors, response format-related factors, generic individual-related factors, or administration factors. For instance, socially desirable responses may be caused by item content- or administration-related factors such as interviewer bias, or a generic individual difference. The resulting error could be additive systematic error, say, interviewer bias or item content, moving responses in one direction. The resulting error could be correlational systematic error, say, item content, interviewer bias, or response format (e.g., end anchors such as *like-hate,* the latter being extreme and perhaps socially undesirable) differentially affecting individuals who differ on tendency toward social desirability. Therefore, consistent differences over and above the construct being measured result. For correlational systematic error to occur, a source has to have a consistent, differential influence across individuals. Again, several of these sources could fit under generic individual differences or under interactions described below.

Sample Interactions

Generic individual-related factors, such as ability or tendency toward impression management, could interact with administration-related factors, such as interviewer bias or item content-related factors, or response format-related factors, such as task-related ability (e.g., computer skills and language processing ability), central tendency, or leniency. Individual differences could be accentuated by response formats. Similarly, item content can interact with generic individual-related factors (such as individual differences in social desirability and item wording to elicit social desirability). Item content can interact with response format through the tone of the item (e.g., extreme wording) and the use of extremes versus the center of a scale. Similarly, administration-related factors can interact with other categories of factors (e.g., interviewer bias and generic individual differences in impression management, response format, or item wording). Administration-related factors, item content-related factors, and response format-related factors can also interact with idiosyncratic individual factors to lead to idiosyncratic random error.

Summary

Many sources can cause each of the types of measurement error described in Chapter 2. By understanding what causes error, these sources can be minimized in the design of items and measures. These sources can be roughly categorized into individual-related sources of errors (idiosyncratic and generic) and method-related sources of errors, the latter being separated into item content, response format, and administration issues. Hence, there are certain characteristics (enduring and transient) that individuals bring into a data collection event, and there are certain characteristics of the data collection event. In turn, the sources in each category can cause the different types of measurement error described in Chapter 2. Although the taxonomy simplifies reality by categorizing sources into single categories, many sources can be categorized as interactions among these categories.

Note

1. The term *method* has been used in different ways covering narrow to broad issues. Method is used here in its broadest sense, to refer to everything that is done to collect the data, including the setting, the administration procedures, and the measure itself. All of the things that are done to collect data can cause error, and all of them are included here in the notion of a method. The term *method* has been used sometimes in the literature to refer to two different ways of collecting data, such as a paper-and-pencil method versus an observation, or a Likert approach versus a behavioral inventory. These are narrower uses of the term because other issues beyond the use of one format versus the other are involved in a method and could cause error.