ATTITUDE MEASUREMENT IN SURVEY RESEARCH

Survey research

Survey research is a type of quantitative research that aims to obtain information about the attitudes or opinions that people have toward the particular issues. However, survey research should not be confused with "*poll*". Although both of them aim to obtain information about attitudes and opinions of people, there are some major differences between them. First, survey research is normally based on sound theoretical background; but for poll, it does not require a theory. Second, survey research aims to study the relationship between phenomena; and thus, the questions being asked in the survey normally comprise a set of related information. In contrast, poll can has only one single question. Third, the results from survey research require systematic data analysis and the findings are normally used to guide decisions. However, the main objective of poll is to capture a snapshot of how people think or feel about a specific issue; there is no policy implication for the poll result. The major differences between survey research and poll are summarized in the table below.

	Survey research	Poll
Theory-based	Yes	No
Data collected	Several related questions	Can be a single question
Objectives	Final result will be used for decision making or policy implication.	To capture just a snapshot of how people think about a specific issue; no policy implication associates with a result.

Understand attitudes

As mentioned earlier that the main objective of survey research is to study about attitudes of people, so we need to have a basic understanding first about the importance of attitudes. *Attitude* is defined as the degree to which a person has a favorable or unfavorable evaluation or appraisal of something (e.g., people, place, objects, events, etc.). In particular, this definition suggests that attitudes normally have two opposite signs: positive (favorable) and negative (unfavorable). If we like something, we will develop positive (favorable) attitude towards that thing. On the other hand, if we dislike something, we will develop negative (unfavorable)

attitude towards it. By the way, sometime attitudes can also be neutral. In this case, people are indifferent about the particular issue; they are neither positive nor negative about it.

Attitudes can be classified into three aspects: cognitive aspect (how we think about something), affective aspect (how we feel about something), and behavioral aspect (our intention to do something). Although these three aspects of attitude reside in different domains, they tend to be interconnected. In particular, research in psychology has shown that our emotion and cognition tend to affect each other (Storbeck & Clore, 2007). Moreover, the interconnection between emotion and cognition can eventually be transformed into specific behaviors that we express (Schwarz, 2000). For example, let's assume that there is one person who always treats you badly. Of course, when someone is not nice to you, it is more likely that you might develop negative attitude toward him. At this point, you may think that you don't like him. Your negative thought that arises in the first place can then be developed into negative feeling. In this case, the fact that you don't like this person will probably make you feel that you hate him as well. Eventually, the negative thought and the negative feeling that you have developed toward that person can motivate you to take some action toward that person also. For example, you will try to stay away from him or to avoid seeing him at all costs.

In academic field, a prominent theory that can explain this role of attitudes on behavioral intention of people is known as *the theory of planned behavior (TPB)*. TPB has been widely used in marketing and information systems research as a theoretical support to explain the influential role of attitude that can guide behavioral intention of people toward doing something (Ajzen, 1985, 1991). The theory was firstly introduced by Icek Ajzen in 1985. The gist of this theory is that there are a linkage between personal belief and behavior. The theory predicts that our intention to do something is driven by three things: (1) attitude toward the behavior, (2) subjective norms, and (3) perceived behavioral control. Anyway, among these three factors, attitude toward the behavior tends to play an inferential role in predicting the actions of people. For example, if we have favorable attitude toward a particular product (let's say we saw a cool mobile phone that we really like), it is more likely that we will buy it whenever we have a chance (and enough money). TPB has been tested in many contexts such as in health-related behaviors

(Godin & Kok, 1996; McEachan et al., 2011), consumer decision toward product/service purchase (Casaló et al., 2010; Han et al., 2010), and technology adoption (Baker & White, 2010; Yousafzai et al., 2010). Overall, results from various areas of research tend to support its predictive power that attitude will lead to actions.

Due to the power of attitudes that can influence our actions, understanding about attitudes can help social scientists predict the behaviors of people that may follow when they develop a particular attitude. In marketing research, for example, the questions that marketers frequently ask in the market survey normally involve (1) the attitudes that consumers have toward a particular product; and (2) the degree to which consumers plan to purchase the product in a future. Marketers believe that the positive attitude that consumers have toward the product can be a key indicator to justify whether the product will have a potential to attract consumers or not. In addition, the degree to which consumers report that they will buy the product in a future can also be used as the indicator to predict the actual purchase of consumers as well.

Measuring attitudes

Basically, attitudes of people are quite difficult to be measured objectively. We cannot measure attitude of people by direct observation. For example, if we want to know the degree to which employees are satisfied with their job, what would you do to measure it? Can you just observe their happy or miserable faces while working to evaluate whether employees are satisfied or dissatisfied with job; then you give job satisfaction score for each employee based on what you observe from their faces? While someone may agree that it makes sense to do so, in practice, this method of attitude evaluation can be highly susceptible to subjective bias from the observer. Furthermore, it is possible that what people reveal through facial expression may not consistent with their inner feelings, thereby making direct observation of attitudes become misleading. For this reason, attitudes are regarded in research as the *hypothetical constructs* that cannot be observed or measured directly. Because of this, the measure of attitude in research is usually represented in terms of the "*latent variable*". In particular, a latent variable is a variable that measure a concept that cannot be directly observed such as thoughts and opinions.

In fact, it has been suggested that the best way to gain access to attitude of people is by asking them to express their attitude openly. Therefore, using qualitative method to collect the attitude data might provide more advantage than using quantitative method. However, this does not mean that we cannot use quantitative method to measure attitudes of people. For quantitative research, the method that is widely used by social scientists to measure attitude is asking people to express their attitude through a rating scale. One particular type of rating scale that is popular in survey research is Likert scaling method.

Likert scaling

Likert scaling is a bipolar scaling method that measures either positive or negative response to a question statement. Likert scaling starts with the question statement that represents the attitude that the researchers aim to measure. The question statement is followed by Likert items. A Likert item is simply an option that the respondent will select to evaluate the question statement being asked. A Likert items are ranked orderly from low to high or from negative to positive. For example:

Overall, I am satisfied with my job.

1. *Strongly disagree*
2. *Disagree*
3. *Neither agree nor disagree*
4. *Agree*
5. *Strongly agree*

The number of options can be odd or even. For Likert scaling with odd number items, respondents are provided an option to answer "neither agree nor disagree" or to be neutral toward the statement being asked. On the other hand, Likert scaling with even number items will force the respondents to either agree or disagree with the statement being asked; they are not allowed to be neutral. For this reason, Likert scaling with even number Likert items is also called a *forced choice method*.

For example:

Overall, I am satisfied with my job.

1. *Strongly disagree*
2. *Disagree*
3. *Agree*
4. *Strongly agree*

Generally, the decision to provide odd or even Likert items depends on the characteristics of the respondents. In the Western culture, odd number Likert items tend to be widely used. However, in the Eastern cultures like in China and Japan, forced choice method tends to be more preferred as research has showed that people in these cultures tend to answer the midpoint option on the scale more frequently than people in the Western culture (Chen et al., 1995; Lee et al., 2002).

Note that the use of Likert item does not limit to the agree/disagree option. They can also be expressed in various terms such as frequency of occurrence, satisfaction, magnitude, etc.

How often do you feel nervous?

1. *Never*
2. *Rarely*
3. *Sometimes*
4. *Most of the Time*
5. *Always*

To what extent are you satisfied with your monthly salary?

1. *Very Dissatisfied*
2. *Dissatisfied*
3. *Neutral*
4. *Satisfied*
5. *Very Satisfied*

COMING UP WITH THE QUESTIONS TO MEASURE ATTITUDE

To come up with the question statement that you will use to measure the attitude of people, there are three methods that you can follow. These methods include:

1. Using the scale that was already developed by other scholars.
2. Adapting the scale that was already developed by other scholars.
3. Developing your own scale.

Generally, the decision to choose which method to setup the questions depends on the availability of the existing scale and whether the existing scale you found can be applied to the respondents that you target in you research. Anyway, each method tends to have its own advantages and disadvantages. These issues will be discussed below.

Using the scale that was already developed by others

The first method, which is using pre-existing scale developed by other scholars, is the method that is quite popular and is widely accepted in academic research. In journal articles, many scholars chose to use the scale that was previously used in other journals to measure their concept because it was already validated by the scholars who developed the scale.

Anyway, how can you know that there are pre-existing scales that you can use and how you can obtain them? Well, it may not be quite easy; albeit not too difficult. In order to know whether the concept you want to measure already has the scale already developed, you have to search the empirical papers that involves that concept. The main reason is because scholars are usually required to report the questions that they used to measure the concepts in the paper. The information about the scales is normally reported under the methods section. In some paper, the authors may provide only the reference to where they got the scales from; some authors may provide just a few sample scale items; but if you are lucky, you may find a paper that even provides the compete questionnaire that they used to collect the data.

Adapting the scale that was already developed by others

Even though you could find the pre-existing scale that you may use to measure your concepts, sometimes the scale that you found may not precisely match with your research context or the characteristics of your samples. In fact, when you obtain the pre-existing scale from other literature, you have to carefully read the question items first to ensure that they can be applied to your samples before you decide to use them.

There are some reasons why you may have to adapt the questions in the scale that you found in previous research. Firstly, adaptation may be required when the research context of the previous study where you got the scale from is different from your research context. Another reason why you have to modify the question items is because sometime the questions developed by other scholars are cultural bounded. Obviously, many measurement scales that have been widely used in research were developed by scholars from the Western countries. Due to culture differences between the East and the West, sometimes what means one thing in the Western culture may not always mean the same thing in the Eastern culture. In this case, you may need to make some adjustment to the question items to make them applicable to the characteristics of people in your culture.

There are many ways that you can adapt the pre-existing scale. Adaptation can be minor or major. For some instance, you may modify the wordings of some question, add more questions, or even remove some questions that do not application for your study. Modifying the questions is ok to perform in academic research; but still, you will need to report in the paper that the scale you use is adapted from the scale originally developed by whom.

Developing your own scale

In some case, it is possible that you cannot find any pre-existing scale that can be used to measure your concept. This is quite normal when your concept is new in research and no one has developed the measurement for it yet. In this case you can develop your own questions to measure the concept.

Still, developing your own scale is not an easy task. In order to develop a good scale, it is important that you need to review the literature related to that concept very well to make sure that the question items you initiate cover key attributes of that concept. In addition, it is important for you to pretest the scale before you use them for large-scale data collection to avoid measurement error.

Using pre-existing scale VS Developing your own scale

	Using existing scale	*Developing your own scale*
Advantages	The measurement is consistent with previous research. The scale was already validated.	It can be useful when there is no pre-existing scale to measure the concept. Questions can be customized to match the research context and/or characteristics of the sample.
Disadvantages	The scale may not be applicable to the research context and/or the characteristics of the sample.	Extensive literature review is required. Reliability and validity can be a major concern. The scale need to be pretested before actual data collection.

Single-item scale VS Multiple-item scale

When you come up with the scale to measure your concept, you have the choice whether you want to use a single-item scale or a multiple-item scale. The main difference between a single-item scale and a multiple-item scale is obvious. For a single-item scale, you only have one question to measure your concept; but for a multiple-item scale, you have more than one question to measure your concept.

The main criterion to justify whether a single-items scale or a multiple-item scale should be used depends on the concept that you want to measure. Generally, a single-item scale can be used when the concept can be clearly understood and can be measured from a single aspect. Some scholars also proposed that the use of single-item measures is acceptable when the concept being measured is narrow or unambiguous to the respondents (Lowery et al., 2002; Wanous et al., 1997).

Although using a single-item scale can be easy and straightforward, it also has some major drawbacks. For example, if the respondents misunderstand the question, they will have higher tendency to give you the wrong answer which will lead to measurement bias. Because there is only one question that measures the concept, the validity of the result will be compromised if the question is answered wrongly. In addition, using a single-item scale is not recommended for the concept that encompasses many aspects and cannot be covered by a single question. For this reason, a single-item scale is not suitable for measuring attitudes of people that are quite abstract by nature.

In particular, using a multiple-item scale can provide more advantages than using a single-item scale for several reasons. First, it helps mitigate the issue that the respondents may misunderstand the question. Using a set of related questions to measure the same concept instead of using a single question will help the respondent obtain some clue about what is being measured by recognizing the similarity in meanings of the question statements that aim to measure the same thing. Furthermore, having multiple related questions for the concept allows the researchers to check whether the respondent understand the concept being asked or not by recognizing the pattern of the answers that the respondent provides.

The example of multiple item scale that serves this objective is the measurement of "job autonomy". In particular, this concept reflects the level of freedom that employees have to perform their work. The scale comprises three questions as the following:

- I have sufficient authority to fulfill my job responsibilities
- I have enough freedom over how I do my job
- I have enough authority to make decisions necessary to provide quality treatment

When you look at these three questions, you may notice that they tend to have similar meaning. Anyway, have you wondered why we need to have three questions that ask something similar; why can't we have just one question for this concept. You may think that it does not make sense. Actually, the purpose of having multiple questions for this concept is to make sure that we can capture the consistency in the answers that the respondent will provide. By repeating the questions that have similar meaning, the respondent who actually has high level of job autonomy at work tend to answer all three questions in the same direction, that is, he/she is more likely to agree on all question statements. In contrast, the respondent who actually experienced low autonomy at work is more likely to disagree on all question statements. The consistency in the answers can be the indicator that all three questions accurately capture the level of job autonomy that the respondent actually has. On the other hand, if the respondent strongly agrees on the first question but strongly disagrees on the second question and the third question, you can suspect that there must be something wrong with this measurement because all three questions are supposed to capture the same concept. In this sense, the purpose of using multiple-item scale is to ensure the validity of the measurement scale. The topic about scale validity will be discussed in detail later.

Second, using multiple-item scale allows researcher to capture many aspects of the concept which cannot be measured by a single question. For example, when measuring the concept "job demands" which represent the characteristics of job that create high quantitative workload to employees (Karasek et al., 1998), using a single-item measurement such as "to what extent do you feel that your job is demanding" may not be sufficient to cover this concept. If you review literature related to this concept, you will find that high demanding job normally comprises several aspects including: (a) work fast, (b) work hard, (c) excessive work, (d) not enough time, and (e) conflicting job demand (Karasek et al., 1998). In particular, the job that is considered real demanding usually incorporate these five aspects of workload. Furthermore, when you consider these five aspects of workload carefully, you can see that they tend to relate highly to one another as well. For instance, you will not have enough time when you have to work hard and have excessive work. Therefore, using only a single question to measure job demands

will prevent us to capture different aspect of workloads that employees might experience in their job.

However, using multiple-item scale also has some disadvantage. The major disadvantage is that more questions that you add to measure a single concept can put more effort to the respondents to answer the survey. Thus, the amount of questions to be used for the multiple-item scale is the issue that you may have to consider. Normally, approximately three to five questions can be considered an optimal number of questions per one concept. However, there is no fix rule regarding how many questions you should have per concept. In academic research, you can see that some concept is measured using two questions while some concept is measured by more than ten questions.

Summated scale

Although the objective of using multiple item scale is to capture different aspects of the concept, the answers that the respondent gave to all questions that belong to the same concept will eventually to be combined into a single numerical value that represents that concept. The *summated scale* can be easily understood as the average score of all answers that belong to a particular concept.

Let's consider three question statements that measure job autonomy. The Likert items range from 1: strongly disagree, 2: disagree, 3: neutral, 4: agree, and 5: strongly agree. The answers that one respondent provided are the following:

- I have sufficient authority to fulfill my job responsibilities
  - Answer 4
- I have enough freedom over how I do my job
  - Answer 5
- I have enough authority to make decisions necessary to provide quality treatment
  - Answer 5

In this case, the summated scale of this concept for this particular individual can be calculated by using the average of all values, which is (4+5+5) ÷ 3 = 4.67.

SOME TECHNIQUE IN SCALING THE CONCEPTS

Reverse scaling

When using the multiple-item scale, some scholars suggest that you should incorporate some reverse coded question as well. In particular, reverse scaling is the technique that is used to detect whether the respondents carefully read the question statements when the answer the survey or not. Before explaining in detail what is reverse scaling, let's take a look at the following question statements:

Please rate the extent do you feel about your "supervisor"
(1: strongly disagrss; 5: strongly agree)

1. My work supervisor really cares about my wellbeing
2. My supervisor cares about my opinions
3. My supervisor shows very little concern for me
4. My supervisor strongly considers my goals and value

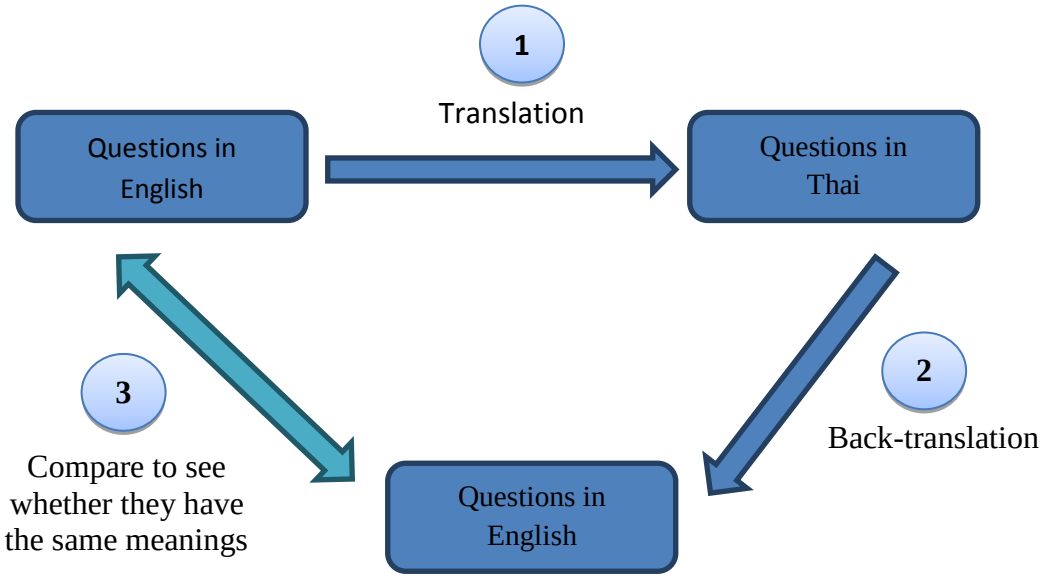
If you read and compare these four questions carefully, you will detect that the meaning of the question number three is obviously different from other questions. While other questions portray about relationship with supervisor in a positive sense, the relationship portrayed in the question number three is negative.

By using reverse scaling technique, the question sentence is worded to alter the meaning from positive to negative or from negative to positive. Generally, reverse scaling is considered a good technique that helps you detect whether the respondents just quickly answer your survey without carefully reading the questions or not. For example, by considering the scale that measures supervisor support (as shown above), if the respondent think that his/her supervisor is supportive, he/she will agree with the question number one, two, and four; but will disagree with the question number three. On the other hand, if the respondent thinks that his/her supervisor is not supportive, he/she will disagree with the question number one, two, and four; but will agree with the question number three. Therefore, if you notice that there is a survey in which the respondent agrees (or disagrees) on all four questions, you can suspect that the responses you get from that survey might be questionable; thus, you may consider removing that survey from the analysis.

Back-translation

Back-translation is the technique that is necessary when you conduct a cross-cultural research that need to translate the question statements from one language to another (Hult et al., 2008). For example, when you obtain the scale that was originally developed in English and plan to use it to collect the data from the respondents whom English is not their native language, you need to translate the questions from English into their native language in order to avoid misunderstanding. However, just translating the questions from English to another language may be problematic unless you are a very experienced language translator.

Translation and back-translation process



After you translate the questions from English to another language (as shown in step 1 in the figure above), it is crucial to ask other bilingual professional to back-translate what you already translated into English again (as shown in step 2 in the figure). This process is called "back-translation". Anyway, it does not make sense if you perform the back-translation by yourself. Then you compare the version

that was back-translated to the original English version to see whether the version that was back-translated still retains the same meaning or not (as shown in step 3 in the figure). If the meaning of the question that is back-translated does not match with your first translation, you should revise your original translation and then resume the back-translation process again until their meanings correctly match.

RELIABLY AND VALIDITY OF THE MEASUREMENT SCALE

For the measurement scale to yield the highest accuracy, it is important for the scale to exhibit sufficient level of reliability and validity.

Reliability

Reliability is the overall consistency of a measure. A measure will exhibit high reliability if it produces similar results under consistent conditions. To have better understand about the meaning of reliability, let's consider this example. Based on technical information from Apple Inc., the exact weight of iPhone 5 is 112 grams. If I use the physical scale that I have to measure the weight of iPhones from 100 people, I should get the same weight of 112 grams from all of them (let's assume that all 100 iPhones that I weight are not fake iPhones from Shen Zhen, China). If I use my scale to measure the same thing (which is iPhone 5) over and over again and still get the same weight of 112, I can conclude that my scale is *reliable* because it can obtain *consistent* results no matter how many units I weight. When applying this situation to the reliability of the survey questions, if you use the same questions to measure the attitude of 100 people who have the same opinion about the questions being asked, you should get consistent responses from all of them either.

In statistics, there are several techniques that scholars use to evaluate the level of reliability of the scale. However, the most widely accepted method in academic literature is by using *Cronbach's alpha (α) coefficient* analysis. Detail about how to perform this statistical method will be provided in the later chapter on data analysis using SPSS.

Validity

While reliability deals with consistency, validity concerns about accuracy. *Validity* deals with whether the scale you use can accurately measure the concept that you aim to measure. In order to have a clear understand about validity, let's take a look

at the table below. If you have some marketing background, you can guess that the table lists four attributes of 4Ps in marketing mix.

Question statements for measuring marketing-mix factors

Product design	*Price*
• The product design is unique. • The product design is trendy. • The product design is attractive.	• The product price is reasonable. • The product price worth the quality. • The product is usually on sale.
Distribution	*Advertising*
• It is easy to buy the product. • The product can be bought at many stores. • It is very convenient to buy the product.	• Advertising of this product is attractive. • Advertising of this product is enjoyable to watch. • Advertising of this product is well-produced.

If you are good in marketing theory, you can see that all question statements for *product design*, *distribution*, and *advertising* appear to match precisely with their underlying concepts. However, for the questions that belong to the concept *price*, there is something not quite right. If you study the marketing theory very well, you will see that the question '*the products are usually on sale*' does not belong to the price aspect of the marketing mix. Instead, it is a part of sales promotion. In marketing, discount is considered a temporary reduction in price that the marketers employ to boost short-term sales. On the other hand, the concept of *price* in marketing is the value that is assigned to the product, which is quite stable. From this example, if you have to evaluate the validity of each concept, you can conclude that the measurements of the concept *product design*, *distribution*, and

advertising have sufficient level of validity because all question statements that belong to each of them correctly represent their underlying concept. On the contrary, the measurement of the concept *price* is not yet valid because there is one question statement that does not correctly represent the concept. In this case, you may consider removing the question that does not belong to the concept or coming up with a new question that correctly matches with the concept to ensure the validity of the measure.

Convergent validity VS discriminant validity

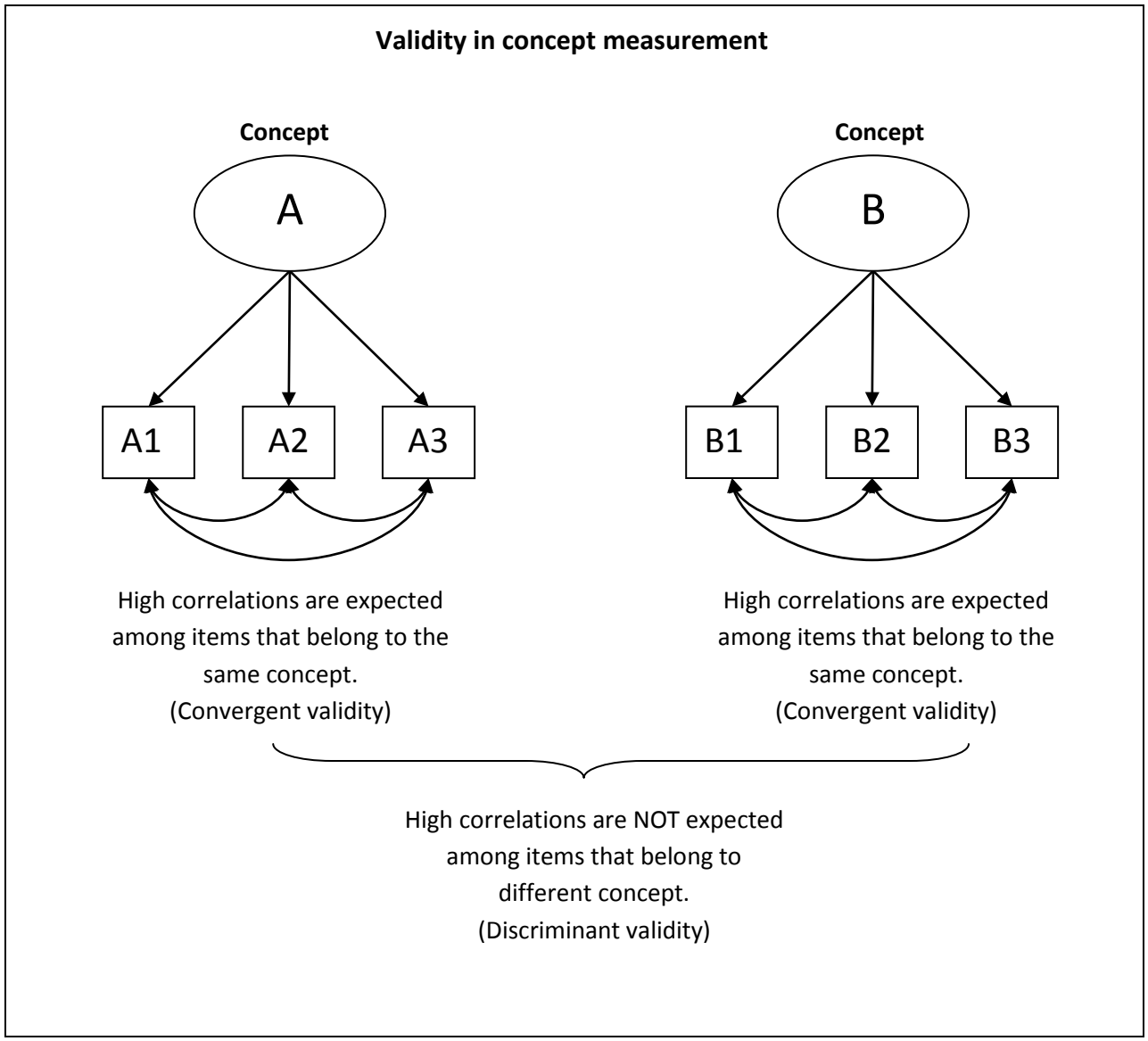
In statistics, validity of the measurement scale can also be classified into (a) convergent validity and (b) discriminant validity. In statistics sense, *convergent validity* is a requirement that all question items that belong to the same concept should share high variation with one another. On the other hand, *discriminant validity* is a requirement that all question items that belong to the same concept should not share high variation with other question items that belong to other concepts.

From the figure above, let's assume that there are 2 concepts named "Concept A" and "Concept B". Concept A is measured by using 3 question items (including A1, A1, and A3); Concept B is also measured by using 3 question items (including B1, B2, and B3). Convergence validity means that all question items that belong to their own underlining concept should correlate highly with one another. If we want to claim that concept A has good convergence validity, we expect high correlations among question A1, A2, and A3. In this regard, high correlation among the questions under the same concept means that the respondents are expected to agree (or disagree) with A1, A2, and A3 in the same direction when answering the survey. In statistical analysis, the level of convergence validity can be evaluated by using the analysis called *factor analysis*.

On the other hand, for discriminant validity to be satisfactory, we expect no significant correlations among the question items across concepts. In this regard, the question items that belong to concept A (including A1, A2, and A3) should not correlate highly with the question items that belong to concept B (including B1,

B2, and B3). In statistical analysis, discriminant validity is evaluated using the analysis of the *average variance extracted (AVE).*

In particular, convergent validity and discriminant validity are the technique required in advanced statistical analysis called *structural equation modeling (SEM)*.



BIAS IN ATTITUDE MEASUREMENT

As mentioned earlier, attitude of human is inherently complicated and cannot be measured objectively in practice. For this reason, the data involve the attitude of people can be highly susceptible to measurement error. *Measurement error* is the difference between the observed value of a phenomenon, which is measured using survey data, and the true value of that phenomenon, which is often impossible to measure (Groves, 2004). Measurement error happens when the concept that is measured does not reflect its true value. Generally, response bias is a common problem in attitude measurement that leads to measurement error. *Response bias* happens when the information that respondents provide are distorted from a truth. Response bias can be intentionally and unintentionally caused by a respondent.

Some types of response bias that are very common in attitude measurement will be discussed as the following:

Social desirability bias

Social desirability bias is "a tendency of research subjects to give socially desirable responses instead of choosing responses that are reflective of their true feelings" (Grimm, 2010). In other words, it is the tendency of people to intentionally distort the information that they provide in the way that makes them look favorable by others. Nederhof (1985) suggested that there are two aspects of social desirability bias (1) self-deception - individuals distort information in the way that enhance their self-esteem; and (2) other-deception - individuals distort information in the way that make them look good by others. Basically, individuals are more likely to overstate activities or characteristics that are socially or culturally desirable, but tend to understate activities or characteristics that are socially or culturally undesirable (Zerbe & Paulhus, 1987). For example, because wealth is perceived favorably in a society, when asking people to report their income, a person may report the amount that is higher than what they actually earned. Sometime people inflate the answers related to their intellectual achievement and performance in order to enhance their self-worth or to conceal their true weaknesses.

Moreover, the chance of social desirability bias tends to be more pronounce when the respondents are asked about the threatening questions (Bradburn & Sudman, 1974). Threatening questions generally involve sensitive issues such as sex, drug, unethical behaviors, etc. When the respondents encounter with these types of questions, they may decide to give social desirable answers or choose to skip answering the questions (Johnson & Delamater, 1976). For example, when asking people to report the frequency that they committed counterproductive work behaviors (e.g., using office equipment for personal benefit or sabotaging coworkers), some people who actually committed these activities may avoid reporting the truth or underreport the frequency of those activities as they deems unacceptable in a society. Sometime people are afraid to report the truth to because they are afraid of the consequences that may follow, or because they feel embarrassed of telling the truth.

Social desirability bias is a common issue that normally happens when self-reported questionnaire survey is used to collect the data (Arnold & Feldman, 1981; Fisher & Tellis, 1998). However, scholars suggested that the chance of social desirability bias in the survey response can also depend on personality differences and national culture of people (Ones et al., 1996). For example, people from Japan tend to avoid expressing negative feelings to others in public (Barrett et al., 2011). For this reason, the Japanese respondents may choose not to report unfavorable attitudes in the survey, even though they are negative about the questions being asked. Interestingly, Bernardi (2006) collected data from students in 12 countries (including Australia, Canada, China, Colombia, Ecuador, Hong Kong, Ireland, Japan, Nepal, South Africa, Spain, and the United States) and found that social desirability bias appeared to be more pronounce for people from collectivism cultures (e.g., China) and high uncertainty avoidance cultures (e.g., Japan) than people from individualism cultures (e.g., United States) and low uncertainty avoidance cultures (e.g., Ireland). Also, they found that the tendency of social desirability bias appeared to be stronger in female than male.

Social desirability bias can be a serious issue that compromises the validity of a concept measurement, and that can subsequently causes bias the results obtain from data analysis (Norwood & Lusk, 2011). Due to this problem, there are

several methods suggested to detect and minimize the chance of social desirability bias in a survey.

Image management subscale (Paulhus, 1986)

1. Sometimes I tell lies if I have to.
2. I never cover up my mistakes.
3. There have been occasions when I have taken advantage of someone.
4. I never swear.
5. I sometimes try to get even rather than forgive and forget.
6. I always obey laws, even if I'm unlikely to get caught.
7. I have said something bad about a friend behind his/her back.
8. When I hear people talking privately, I avoid listening.
9. I have received too much change from a salesperson without telling him or her.
10. I always declare everything at customs.
11. When I was young, I sometimes stole things.
12. I have never dropped litter on the street.
13. I sometimes drive faster than the speed limit.
14. I never read sexy books or magazines.
15. I have done things that I don't tell other people about.
16. I never take things that don't belong to me.
17. I have taken sick leave from work or school even though I wasn't really sick.
18. I have never damaged a library book or store merchandise without reporting it.
19. I have some pretty awful habits.
20. I don't gossip about other people's business.

Detecting social desirability bias

Scholars proposed that some scales can be used to detect social desirability bias in the questionnaire survey. These scales include the Marlow–Crowe scale (Crowne & Marlowe, 1960) and the Balanced Inventory of Desirable Behavior scale (Paulhus, 1986). Some question statements used in the Balanced Inventory of Desirable Behavior are shown in the table above. These instruments can be used to

capture the level of social desirability bias that may inherit in the survey response. In particular, high score of the overall social desirability bias scale can be a sign of this problem. Accordingly, some scholar suggested rejecting the data of high-scoring subjects to avoid bias in the results (McGuire, 1969).

Preventing social desirability bias

The solutions to avoid social desirability bias in self-report measure can be performed in different steps of questionnaire design. Before presenting the questions to the respondents, it is essential for the researchers to provide clear information on the purpose and rationale of the research. This information can be placed in the cover letter. It should provide clearly understandable and easily verifiable procedures that reduce potential embarrassment and ensure confidentiality (Gregson et al., 2002). In addition, some studies suggested that allowing anonymity in the survey response can potentially reduce the chance of social desirability bias in the survey (Joinson, 1999).

Direct question

You stole office equipment from your workplace whenever you had a chance.

Never (1) (2) (3) (4) (5) Always

Indirect question

Stealing office equipment from a workplace is acceptable.

Strongly disagree (1) (2) (3) (4) (5) Strongly agree

Moreover, when dealing with the threatening questions that might trigger social desirability bias, Fisher (1993) proposed that using indirect questioning can be a method to overcome the social desirability bias problem. In particular, Arnold and

Ponemon (1991) suggested that asking the questions in the third person perspective can serve as one powerful solution.

Generally, asking the questions in the third person perspective can be performed by simply asking the respondents about what other people think regarding some potentially sensitive issue. This method can provide a reliable measure of what the individual actually believes. The main objective of this technique is to reduce the distortion of private opinions that respondents reveal to the researcher. According to Fisher and Tellis (1998), "this technique rests on the assumption that respondents project their unconscious biases into ambiguous response situations and reveal their true feelings about socially-sensitive issues". By using this method, the respondents can feel that they are providing information about the situations based on fact rather than opinion (Simon & Simon, 1975), thereby making them more comfortable to express their own opinions and attitudes to the questions openly.

Acquiescence bias

Another type of respondent error is known as acquiescence bias. *Acquiescence bias* (also be known as "yea saying") is the tendency that the respondents choose to agree with all questions in the survey. Acquiescence bias is common when the respondents are asked to assess the question statement using agree-disagree Likert items. Generally, the source of acquiescence bias can come from several reasons. For example, it can happen when the respondents are being friendly when answering the survey; and thus, they may feel that they have to agree on the questions being asked in order to satisfy the researchers, not answer from their true opinion. Acquiescence bias can possibly happen when the respondents just quickly agree on all question statements in the survey without paying attention to the questions. In research, acquiescence bias is a serious issue in attitude measurement because the answers that the researchers obtain in the survey do not reflect the true attitude that the respondent has toward the issue being asked (Podsakoff et al., 2003).

Research has suggested some solution to detect the presence of acquiescence bias in the survey data. For example, reverse-coded question can be used to detect

whether the respondents just simply agree on all questions or not. If the sign of acquiescence bias is detected in any survey, that survey should be removed from the dataset in order to avoid further bias in the results.

Unconscious representation

Unconscious representation happens when the respondents are willing to provide the truthful answers, but for some reasons, they do not aware that the information they provide to researchers is incorrect. There are several reasons that cause unconscious representation. For example, the questions being asked may not be stated clearly. Sometime, the respondents may misunderstand the questions. Moreover, when respondents were asked about activities that happened in the past, they may not correctly recall information that happen long time ago. It is also possible that some questions may be too complicated and is difficult to be recalled. Some example of the question that may susceptible to unconscious representation is asking the respondents how many times that they have checked their Facebook per day. Because not so many people are able to recall the exact information about this, the answer provided for this question may not accurately represent the actual frequency that they accessed to Facebook.

Administrative error

Measurement error in attitude measurement not only arises only from a respondent side, they can also come from people who administer the data collection. This source of bias is known as an *administrative error*. For example, the researcher may accidentally record or code the data wrongly. Sometime, administrative error can happen intentionally when the research cheat on the data. Whether it occur intentionally or intentionally, these are serious issues that researchers must not let them happen.

Interviewer error (or interviewer effect)

One major type of administrative error that commonly happen is the error generated from the interviewer. This type of error is called interviewer error. *Interviewer error* happens when "data collected by either a specific individual

interviewer or a specific set of interviewers may be different than data collected by another individual or set of interviewers administering the same questionnaire to a sample from the same population of respondents" (Davis et al., 2010, p. 15). Interviewer error is also known as *interviewer effects*, which refers to measurement error attributable to a specific interviewer demographic characteristic such as gender (Dijkstra, 1983). Interviewer error is the problem that is quite common is structured face-to-face interviews. In particular, the presence of the interviewer as well as his/her behaviors when administering the survey can potentially influence the answers that the respondent will provide. For example, the respondents may feel uncomfortable to discuss some issues frankly with the interviewer because their identity is not concealed. For this reason, social desirability bias can easily happen. Like respondent error, interviewer error can potentially lower the validity of the survey results that the researchers obtain from data analysis (O'Muircheartaigh & Campanelli, 1998).

Generally, people tend to provide different answers when they are asked by different interviewers for several reasons. For example, research has shown that respondents are more comfortable to provide information to the interviewers who have similar sociodemographic characteristics like them, rather than the interviewers who are different from them (Lenski & Leggett, 1960). Furthermore, some research has documented the role of gender on interviewer effects. For example, Landis et al. (1973) reported that female respondents tended to express more feminist responses to a male interviewer than to a female interviewer. The study by Kane and Macaulay (1993) on gender inequality found that both male and female respondents tended to express more egalitarian gender-related attitudes or greater criticism of existing gender inequalities to female interviewers.

Literature suggests some solutions to deal with interviewer effects. For example, scholars suggested that the chance of interviewer effects can be minimized when training procedures are properly provided to the interviewers (Davis et al., 2010). In particular, good interpersonal skills and trust building can help the respondents feel more comfortable to discuss the issues openly with the interviewers. In addition, Catania et al. (1996) recommended that allowing respondents to select the gender of their interviewers can increase the quality of the information obtained in the interview.

REFERENCES

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control* (pp. 11-39): Springer Berlin Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Arnold, D., & Ponemon, L. (1991). Internal auditors' perceptions of whistle-blowing and the influence of moral reasoning: An experiment. *Auditing: A Journal of Practice and Theory, 10*, 147–165.

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal, 24*(2), 377-385.

Baker, R. K., & White, K. M. (2010). Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective. *Computers in Human Behavior, 26*(6), 1591-1597.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286-290.

Bernardi, R. (2006). Associations between hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics, 65*(1), 43-53.

Bradburn, N. M., & Sudman, S. (1974). *Response effects in surveys*. Chicago, IL: Aldine.

Casaló, L. V., Flavián, C., & Guinalíu, M. (2010). Determinants of the intention to participate in firm-hosted online travel communities and effects on consumer behavioral intentions. *Tourism Management, 31*(6), 898-911.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly, 60*(3), 345-375.

Chen, C., Lee, S.-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among east asian and north american students. *Psychological Science, 6*(3), 170-175.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology 24*(4), 349-354.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research, 25*(1), 14-26.

Dijkstra, W. (1983). How interviewer variance can bias the results of research on interviewer effects. *Quality and Quantity, 17*(3), 179-187.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questions. *Journal of Consumer Research, 20*(September), 303-315.

Fisher, R. J., & Tellis, G. J. (1998). Removing social desirability bias with indirect questioning: Is the cure worse than the disease? In J. W. Alba & J. W. Hutchinson (Eds.), *Advances in consumer research* (pp. 563-567). Provo, UT: Association for Consumer Research.

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion, 11*(2), 87-98.

Gregson, S., Zhuwau, T., Ndlovu, J., & Nyamukapa, C. A. (2002). Methods to reduce social desirability bias in sex surveys in low-development settings: Experience in zimbabwe. *Sexually Transmitted Diseases, 29*(10), 568-575.

Grimm, P. (2010). Social desirability bias *Wiley international encyclopedia of marketing*. London, UK: John Wiley & Sons, Ltd.

Groves, R. M. (2004). *Survey errors and survey costs*. Hoboken, NJ: John Wiley & Sons.

Han, H., Hsu, L.-T., & Sheu, C. (2010). Application of the theory of planned behavior to green hotel choice: Testing the effect of environmental friendly activities. *Tourism Management, 31*(3), 325-334.

Hult, G. T. M., Ketchen, D. J., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Harmancioglu, N., Huang, Y., Talay, M. B., & Cavusgil, S. T. (2008). Data equivalence in cross-cultural international business research: Assessment and guidelines. *Journal of International Business Studies, 39*, 1027–1044.

Johnson, W. T., & Delamater, J. D. (1976). Response effects in sex surveys. *Public Opinion Quarterly, 40*(2), 165-181.

Joinson, A. (1999). Social desirability, anonymity, and internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*(3), 433-438.

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly, 57*(1), 1-28.

Karasek, R., Brisson, C., Kawakami, N., Houtman, I., Bongers, P., & Amick, B. (1998). The job content questionnaire (jcq): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology, 3*(4), 322–355.

Landis, J., R., Sullivan, D., & Sheley, J. (1973). Feminist attitudes as related to sex of the interviewer. *The Pacific Sociological Review, 16*(3), 305-314.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a likert scale. *Research in Nursing & Health, 25*(4), 295-306.

Lenski, G. E., & Leggett, J. C. (1960). Caste, class, and deference in the research interview. *American Journal of Sociology, 65*(5), 463-467.

Lowery, C. M., Beadles, I., N. A., Petty, M. M., Amsler, G. M., & Thompson, J. W. (2002). An empirical examination of a merit bonus plan. *Journal of Managerial Issues, 14*(1), 100–117.

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. *Health Psychology Review, 5*(2), 97-144.

McGuire, W. J. (1969). Suspiciousness of experimenter's intent. In R. R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research*. San Diego, CA: Academic Press.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263-280.

Norwood, F. B., & Lusk, J. L. (2011). Social desirability bias in real, hypothetical, and inferred valuation experiments. *American Journal of Agricultural Economics, 93*(2), 528-534.

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 161*(1), 63-77.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660-679.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleiter & J. S. Wiggins (Eds.), *Personality assessment via questionnaire: Current issues in theory and measurement*. Berlin: Springer-Verlag.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion, 14*(4), 433-440.

Simon, J., & Simon, R. (1975). The effect of money incentives on family size: A hypothetical-question study. *Public Opinion Quarterly, 38*(Winter), 585-595.

Storbeck, J., & Clore, G. L. (2007). On the interdependence of cognition and emotion. *Cognition and Emotion, 21*(6), 1212-1237.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: Howgood are single item measures? *Journal of Applied Psychology, 82*, 247–252.

Yousafzai, S. Y., Foxall, G. R., & Pallister, J. G. (2010). Explaining internet banking behavior: Theory of reasoned action, theory of planned behavior, or technology acceptance model? *Journal of Applied Social Psychology, 40*(5), 1172-1202.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Journal, 12*(2), 250–264.