A PILOT STUDY: VITAL METHODOLOGICAL ISSUES

Osama A. HAZZI[1], Issa Sh. MALDAON[2]

Department of Business Administration, Damascus University, Damascus, Syrian Arab Republic
E-mails: [1]osama.hazzi@hotmail.com (corresponding author); [2]issa-ma@scs-net.org

Abstract. A pilot study represents a vital step for conducting a full-fledged study soundly. In fact, a well-conducted pilot study can help the researchers to design a clear road map they can follow. Conducting a pilot study professionally, however, involves vital quantitative methodological issues (i.e., Back-Translation, Missing Data, Normality, and Reliability) the researchers had better take into account when conducting the pilot study and before embarking the main one. Unfortunately, such issues are neglected in many researches as well as barely addressed together in a one work in research papers. From the literature and the experience, the present work aims to address the most recommended practices of such methodological issues providing at the same time the best practices for new researchers and scholars alike, with a view to reduce the loss of suffered by research community when such issues and practices remain unpublished together. Suggestions and thoughts were also demonstrated in this work for further researches. Finally, the importance of such research for business sector was also demonstrated in this work.

Keywords: pilot study, quantitative methodological issues, back-translation, missing data, normality, reliability.

JEL Classification: C18.

Introduction

A pilot study represents a cornerstone of a good research design. In fact, a pilot study is an essential initial step in a research and this applies to all types of research studies. The term of pilot study, however, is defined as "a small-scale test of the methods and procedures to be used on a large scale …" (Porta 2008). On the other hand, there is little published guidance with respect to the sample size required for pilot studies. The study of Billingham *et al.* (2013) mentioned that even though all studies should have a sample size justification, some kinds of studies do not need to have a sample size calculation. Their studies, however, concluded that a formal sample size calculation for pilot studies may not be appropriate. Generally, 10–20% of the main sample size is a reasonable number for conducting a pilot study (Baker 1994).

Mainly, the importance of the pilot study lies in improving the quality and the efficiency of the main study.

Put another way, a pilot study can be used to reveal some logistics issues before embarking the main study, which pilot study results can inform feasibility and identify modifications needed in the main study. There are also other reasons to conduct a pilot study, for example but not limited to, checking the words and statements of the used scales, refining the scales items, developing scales items and research plan, and collecting preliminary data are indeed some examples for conducting a pilot study. In this regard, even though conducting a pilot study provides us with limited information comparison with the main study and does not guarantee success in the latter, but it does increase the likelihood. A one thing the researchers should pay attention that a pilot study is not a hypothesis testing study. Leon *et al.* (2011), however, mentioned that the main purpose of conducting a pilot study is examining the feasibility of the intended approach the researchers will use in the main study. Generally, a pilot study can be used as a small version of a full-scale study or trial run in

preparation for a main study (Polit *et al.* 2001). It can also be used to check out a particular research instrument.

However, the present work believes that addressing a pilot study is an interesting and important topic amongst the researchers. Furthermore, many researchers disregard conducting a pilot study since it includes quantitative methodological issues (i.e., Back-Translation, Missing Data, Normality, and Reliability) which need long time to look at deeply, especially, for the researchers who are not well-established in research methods. Therefore, many of those researchers consider looking at such issues as obstacle itself and may be overwhelming and time-consuming, so they disregard conducting such issues of the pilot study and move to conduct the main study indifferent to the importance of pilot study. The present work, however, believes that discarding such issues will give misleading results, so it strongly recommends that the researchers had better deal with such issues before investing a lot of time, money, and effort in the main study. In fact, the present work believes that conducting a pilot study first will help the researchers to not sticking with the results they cannot change, and to manage then the main study more flexibility.

On the other hand, such of methodological issues are barely addressed together in a one work, especially, for researches which use survey as the main tool to collect data (e.g., Social, Business, Management, and behavioural researches). Therefore, the present work aims to address the most important methodological issues together and trying hard to give the researchers, especially for those who are not well-established in research method and thirsty for such issues, the best recommended practices when conducting a pilot study and before conducting the main one. It is also aims to provide those researchers applicable insights a view to reduce the loss of suffered by research community when such issues remain unpublished in a one work. Further recommendations and suggestions are also demonstrated in this work for those researchers to avoid the research problems they usually face, especially, for those who are using similar methods. Therefore, the present work tried hard to address such methodological issues in an understandable way which helps those researchers to apply easily and to avoid the researcher problems then.

1. A pilot study: be smart and do its parts

1.1. Back-translation: a base for a better case

Data represents the lifeblood of a research. It helps us to understand the real world well through connecting the theory to practice. Therefore, the researchers should handle with data carefully and honestly, especially, when collecting, analyzing and interpretation.

However, there are different ways to collect data (e.g., questionnaire, interviews, observations, diaries and journals). Questionnaire is one of the most widely used data-gathering instruments in many fields including Business, Management, Market Research, Psychology and Sociology. On the other hand, the researchers usually face matter when they want to use an original questionnaire in a different language. In this case, the researchers should translate the questionnaire to the target language professionally. At the same time, we cannot assume that the translated questionnaire or items are simply valid due to the translation process. Therefore, a data collection instrument, not only questionnaire, must be validated. In this situation, Griffee (2012) defined valid that the translated items would be understood in a way similar to the way intended by the questionnaire maker. Based on the experience, the present work strongly agrees with that grasp since it believes that the translation process does not only lie in literally translation, but also in translating the meaning and the intention of the original data collection instrument. Therefore, translation process is an important and critical step the researchers should take into account.

Reviewing the related literature, the present work finds Back-Translation is an effective technique for translation of a data collection instrument. Generally, Back-Translation is considered the first step the researcher should take into account especially when conducting a research of the pilot study or even the main one, and particularly when using original questionnaire in a different language. However, two bilingual experts, who are preferred to be familiar with the content of the material of a data collection instrument, should work on this kind of translation, that one translating from the source to the target language, while the second blindly translating back from the target to the source (Brislin 1970). On this point, that study recommends that if the two source language versions are not identical, a researcher can then consult with the two bilingual experts for revising purpose.

According the experience, the present recommends that the researchers asking two bilingual professors, that the first one, who translating from the source to the target, is a specialist in the source language and particularly familiar with the terminology of the area covered by the instrument, while the second one, who blindly translating back from the target to the source, is a specialist in the material of a data collection instrument itself. Following such way, the present work believes that Back-Translation technique will not only be an effective technique, but also an efficient one. Put another way, following such way will give the researchers better results and reduce revising process, if any. In any case, translation does not represent a short-cut validation solution. Therefore, translation results must also be piloted and analyzed.

1.2. Missing data: better to manage, not to damage

Once the collected data is ready to handle, the researchers often face some issues regarding missing data. In fact, missing data is one of the most important methodological issues a pilot study or even the main one includes. Moreover, without handling the missing data, neither the pilot study nor the main study works well.

However, before addressing the topic of missing data, researchers who usually use a questionnaire as a data collection instrument face issue regarding the items which need to be reversed-scored. Most of the items usually are phrased to have the same oriented, while other items are phrased in the reverse. In this situation, the researchers need to reverse score those few items to make them comparable to the majority of items. For example, in the questionnaire, which uses Likert scale ranging from 1 ("Strongly disagree") to 5 ("Strongly agree"), the researcher wants the items, which need to reverse-scoring, to turn into the appropriate opposite (i.e., 1's into 5, 2's into 4, 3's into 3, 4's into 2, and 5's into 1). The present work, however, recommends that the researchers must reflect the items, which need to be reversed-scored, prior to handling and replacing missing values.

Returning to the missing data, researchers can find that such issue is a common problem in a research. It clearly appears when the researchers use paper-based instrument for collecting data (e.g., questionnaire or survey). The researchers may also face the issue of missing data even when using web-based instrument, as the respondents may be unable to complete the questionnaire or survey due to server crash or a browser freeze, which results missing data. However, Kang (2013) recommends researchers that the best solution to the missing data is to maximize the data collection when the study protocol is designed and the data collected. Then, the researchers should only use the sophisticated statistical techniques after the maximal effort have been employed to reduce the missing data in the design and prevention techniques.

On the other hand, Bennett (2001) mentioned that if the effect of missing data is not taken into account, the results of the statistical analyses will be biased and the amount of variability in the data will not be correctly estimated. In other words, the researcher had better handle missing data before starting in the analyses since they can have a significant effect on the conclusions that can be drawn from the collected data. In addition, the way the researcher deals with missing data relies on some issues (e.g., how much data is missing, the kind of missing data, and the reasons of "missing data").

There is no, however, consensus with respect to the percentage of missing data that becomes problematic. When Schafer (1999) recommends 5% as the cutoff, Bennett (2001) claims that more than 10%, the missing data becomes

problematic, while others recommend 20% as cutoff (Peng *et al.* 2006). Whatever the percentage of missing values is, the present work believes that the most important point the researcher should also take into account is checking the pattern of missing data to see then how they can handle.

According to the reviews, there are three kinds of missing data (i.e., Missing Completely At Random "MCAR", Missing At Random "MAR", and Missing Not At Random "MNAR"). With MCAR data, there are no patterns in the missing data and the missing values are not related to any variables of the study (Bennett 2001). This implies that there are no systematic reasons for missingness. In other words, missing completely at random (MCAR) data is not intentionally process by the respondent. Schafer and Graham (2002), however, claimed that MCAR is an important special case of missing at random (MAR). With Missing At Random (MAR) data "the most frequently seen in practice", the missing data are related to another variable in the data set but is not related to the variable of interest itself (Allison 2001).

However, with missing not at random (MNAR) data, the missing data are related to the values of the variable of interest itself. Unlike missing completely at random (MCAR), missing not at random (MAR) attributed to systematic reasons for missingness. Put another way, missing not at random is intentionally process by the respondent. Missing not at random (MNAR) is the most difficult case to model for. However, the present work strongly recommends that the researchers had better use a Little's MCAR test, an omnibus statistical test, to find out whether or not the data are missing completely at random. In general, if the *P* value for this test is more than 5%, then this indicates the data are missing completely at random (MCAR). Otherwise, the data are either missing at random (MAR) or missing not at random (MNAR). The researcher should note that such test is looking at the data set as a whole, not the individual variables. In addition, the present work also strongly recommends that the researchers had better statistically test the relations between observed variables and missing values through using a dummy variable. In this method, the researcher simply creates a dummy variable for each missing and nonmissing data set. If the dummy variable, the missingness one, is associated to other variables, then the pattern of data is missing at random (MAR). However, applying those two methods together (i.e., Little's MCAR test and the dummy variable) will help the researcher to determine accurately the kind of missing data. Therefore, the present work strongly recommends that the researchers had better apply these two ways.

However, once the researchers determine the kind of the missing data, they should handle them correctly. There are some techniques for handling missing data in a research, which lay in imputation methods (i.e., nonstochastic

imputation methods including: mean imputation, regression imputation, and pattern-matching imputation "i.e., hot-deck and cold-deck imputations"; & stochastic imputation methods including: stochastic regression imputation, Expectation Maximization "EM", Full Information Maximum Likelihood "FIML", and Multiple Imputation "MI"), and deletion methods (i.e., Listwise, Pairwise). Note that such of those methods, except multiple imputation, fall under single imputation group.

It is no secret that deletion methods are generally not recommended for handling missing data; however, imputation methods are alternative methods. On the other hand, not all techniques of the latter are effective. For example, when mean imputation, which assumes that data are missing completely at random (MCAR), produces biased means with data that are missing at random (MAR) or missing not at random (MNAR) and underestimates variance and covariance, regression imputation produces unbiased means with data that MCAR or MAR, but still produces biases in the variances and covariances; while pattern-matching imputation still suffer from problems (Schlomer *et al.* 2010).

From another standpoint, stochastic imputation methods are effective methods for handling missing data. For example, expectation maximization (EM), which is one of maximum likelihood (ML) approaches, provides "unbiased and efficient" parameters (Graham *et al.* 2003). Even though expectation maximization provides unbiased and efficient parameters, but it does not provide standard errors and confidence intervals. Multiple imputation (MI), however, provides accurate standard errors and therefore good for inferential conclusions. While, multiple imputation technique is more comprehensive than expectation maximization, it is the most complex of stochastic imputation method so far (Schlomer *et al.* 2010). Full Information Maximum Likelihood (FILM), however, represents an alternative efficient stochastic imputation technique, as it produces unbiased results and estimates accurate standard error and confidence interval by retaining the sample size as well as it produces similar results to each of EM and MI (Enders, Bandalos 2001; Olinsky *et al.* 2003). From another standpoint, while scholars consider that FIML is an efficient technique and one of the preferred methods for handle missing data (Enders, Bandalos 2001; Schlomer *et al.* 2010), others believe that MI is probably most promising, because of its theoretical and distributional underpinnings, while FIML is a direct model-based method for estimating parameters in the presence of missing data and does not actually impute the missing data (Olinsky *et al.* 2003). However, full information maximum likelihood (FIML) is a superior method comparing to multiple imputation (MI), as its ability to manage missing data and conduct analyses in one-step, as well as it is a superior method comparing to expectation maximization (EM), as its ability to estimate accurate

standard errors and confidence intervals by retaining the sample size. Therefore, FIML has distinct advantages over both of MI and EM (Schlomer *et al.* 2010).

Putting together, knowing each of the percentage of missing data, the kind of such data, and kind of the most techniques for handling missing data including its pros and cons, the present work can provide the researchers generally different points of view for handling missing data. From the conservative point of view, the present work recommends that the researchers should use stochastic single imputation, expectation maximization (EM) technique, when the missing data are missing completely at random (MCAR) or missing at random (MAR) and when the percentage of missing data is less than 5%. On this point, EM is the most preferred technique of single imputation methods, as in the other methods the variance and standard error are reduced and the chance for Type II errors increases. However, when the percentage of missing data more than 5%, the researchers should use then multiple imputation (MI) technique. From another standpoint, Schlomer *et al.* (2010) believe that both of multiple imputation (MI) and full information maximum likelihood (FIML) techniques viable strategies for handling missing data, as they provide acceptable estimations of regression coefficients and standard errors when the data are missing completely at random (MCAR) or missing at random (MAR) and the percentage of missing data is 10% and 20%. Eekhout *et al.* (2013) recommend multiple imputation (MI) when the percentage of missing data more than 25%. However, even though the researcher could use both of multiple imputation (MI) and full information maximum likelihood (FIML) for handling missing not at random (MNAR), there is also a well-known technique when data are missing not at random (MNAR) represents in Heckman's (1979) method (Heckman 1979).

However, the present work invites the interested researchers to refer to other related literature that deeply address techniques for handling missing data (Allison 2001; Schafer, Graham 2002; Olinsky *et al.* 2003; Rubin *et al.* 2007; Buhi *et al.* 2008; Graham 2009; Schlomer *et al.* 2010).

1.3. Normality: unfortunately, the forgotten assumption by many researchers

Many researchers consider that checking normality is not that important issue when conducting a pilot study. The present work, however, believes that checking normality is also one of the most important quantitative methodological issues the researchers should take into account when conducting the pilot and the main study.

It is known that a normal distribution looks like a symmetric bell-shaped curve, and the mean and variance (variability) are the two main keys to define the normal distribution. But, many researchers disregard the fact that the sample size plays a role in the normality. Under statistical

inference, when the sample size increases, the distributions will close to the normal. Krithikadatta (2014) mentions that when the sample size increases to 25, the distribution is beginning to conform the normal curve and becomes normally distributed when sample size is 30. Therefore, the researchers should be careful when determining the normality, especially when conducting a pilot study, since they generally will have a small sample size comparing to the main study.

Generally, the importance of normality lies in that most of the statistical tests, called parametric tests, rely upon the assumption that the data is normal distribution. Such those tests increase the chances of finding significant results. Otherwise, the researchers should use non-parametric tests when the data is not normal distribution. However, there are ways for checking the normality including eyeball tests (e.g., histogram and normal Q-Q plot), descriptive statistics, and conducting certain statistical tests (e.g., Kolmogorov-Smirnov and the Shapiro-Wilk tests).

Warner (2008) mentions that using histogram for checking the normality is a sufficient way in most situations, which such way probably works for sample size from 25 to 30 or more but not for smaller. From the present work's perspective, checking the normality by examining the shape of histogram can cause confusion for many researchers, since in fact there are infinitely different distributions shapes that may be normal (Bump 1991).

However, descriptive statistics (i.e., Mean, Median, Mode, & Skewness and Kurtosis) is also an easy and good way to check the normality. In fact, when the mean, median, and mode are the same or similar, the distribution is normal. In practice, when the mean, median, and mode are close, the distribution is within the normal range. In addition, skewness and kurtosis is also an easy and effective way for checking the normality. As the rule of thumb, the researchers can argue that the distribution is within normal, if skewness and kurtosis within ±2.0 (Cameron 2004). From another standpoint, West *et al.* (1996) proposed a reference, Satorra-Bentler statistics, of substantial departure from normality as an absolute skewness value <2 and an absolute kurtosis 7. According to the experience, the present work, believes that descriptive statistics is an effective solution for checking the normality for both of small and large sample sizes, especially, when using skewness and kurtosis.

On the other hand, using statistical tests such as Kolmogorov-Smirnov and the Shapiro-Wilk tests is another way for checking the normality. While Elliott and Woodward (2007) recommends using Shapiro-Wilk test for only sample size less than 50, Garson (2012) recommends using Shapiro-Wilk test form small to medium sample size up to N = 2000, and using Kolmogorov-Smirnov test for larger sample. However, when a *p*-value is 0.5 or less, the distribution is not normal. From another standpoint, the

researchers should be aware that Kolmogorov-Smirnov and the Shapiro-Wilk tests of normality are not always work well since they are not sensitive enough at low sample size or overly sensitive to large sample sizes and then give us misleading results. The present work strongly tends to agree with that standpoint according to the initial results of on-going work the present paper work on, which its pilot study is 23. The present work, however, invites the researcher to refer to some resources with respect to using Kolmogorov-Smirnov and the Shapiro-Wilk tests for small sample sizes (see Razali, Wah 2011).

However, the present work believes that using more than one way for checking the normality (e.g., Comparing Mean, Median, and Mode; using Skewness and Kurtosis; and examining carefully the histograms) will generally put the researchers on the safe side and particularly for pilot studies purpose. The present work, however, invites the researchers to refer for other sources since there are several ways for checking the normality and the present work cannot mention all in this paper (see Fisher skewness and kurtosis coefficients in Pett 1997; Brown's two standard errors of skewness in Brown 1997; Cramer, Howitt 2004; Garson 2012; Field 2013).

In the context of talking about normality, the present work stresses on two related topics, outliers and data transformation, even though the present work intentionally putted them off to talk about at first. Outliers are extreme values comparing with the rest of the data. Generally, some researchers simply remove the outliers to render their data normal. The present work, however, does not recommend removing the outliers since such topic is subjective. In addition, removing outliers will not result pure data. Moreover, outliers may be in reality valid data (e.g., form simply opinions) in some researches (e.g., social or behavioural research) and the researchers should not easily remove such data without knowing the main reasons behind such outliers. Over and above, removing outliers may be cumbersome and sometimes overwhelming. When using descriptive statistics (i.e., mean, median, and mode) for checking the normality with the presence of outliers, the present work strongly recommends that the researchers compare the mode and the median and not rely on the mean since the outliers have effect on mean.

Data transformation, however, is also another topic the present work stresses. Data transformation means transform the non-normal data to normal. The present work also does not recommend data transformation since such process will result neither pure data nor honest data. From the point of view of the present work, the researchers have some solutions to deal with the non-normal data. Leave the data as it, and conducting non-parametric tests; or conducting robust statistics, which are just as powerful as parametric tests but account for non-normal data. However,

the present work invites the researchers to refer to sources, which addresses such topic (see Hoaglin *et al.* 1983; Tabachnick, Fidell 2007; Howell 2007).

1.4. Reliability: your way to go ahead confidently

One of the most common mistakes quite a few researchers do is using validated questionnaire or survey but skipping the issue of reliability checking. In fact, reliability is an essential issue the researchers should take into account when conducting the pilot or/and the main study even if they use questionnaires which established before.

However, reliability of a measure ensures consistent measurement cross time and across the various items in the instruments (Bajpai, S. R., Bajpai, R. C. 2014). Deeply, reliability can be expressed in terms of the stability and the consistency (homogeneity), which the instrument measures the concept and helps to assess the goodness of a measure (Sekaran, Bougie 2010).

The method of Test-Retest Reliability is a way the researcher could use to ensure consistent measurement across time. According to such method, the researcher should repeat the same measure on the same respondents after some time (e.g., few days, weeks or even months). Parallel-Form Reliability is also other method to ensure the stability of a measure over time. According to such method, the researchers should use two comparable sets of measures addressing the same construct, but including different order or sequence of questions or sometimes including changing wording of items at second time.

On the other hand, the consistency, which indicates to the homogeneity of items in the scale that measure the same construct, could be examined through some methods (e.g., Inter-Item Consistency and Split-Half Reliability). Cronbach's alpha and Kuder-Richardson Formula 20 (KR-20) are ways of Inter-Item Consistency. However, the present work will focus on Cronbach's alpha since such method can be used for both binary-type and large-scale data, while KR-20 can be used only for dichotomous response scales (i.e., True/False; Yes/No). On this point, the present work invites the researchers to refer to the related source for such purpose (see Kuder, Richardson 1937). On the other hand, while Split-Half Reliability reflects the correlations between two halves of an instrument, the two subsets of items which already were as a one single test, Cronbach's alpha reflects the mean of all the possible split-half reliability estimates of an instrument. As a result, Cronbach's alpha is highly recommended over these two methods.

Generally, Cronbach's alpha is the most common measure for checking reliability (Field 2005). In fact, the researchers can find, through reviewing the literatures, that Cronbach's is the most commonly used in many disciplines, especially, when studies addresses multiple Likert-type scales in Business, industrial and social psychological research.

However, Cronbach's alpha is most appropriately used when the items measure different substantive area within a single construct (Tavakol, Dennick 2011). In other words, Cronbach's alpha is a measure of the internal consistency of a scale, which describes the degree to which all the items of scale measure the same construct. Deeply, it is assumed that alpha test should minimize the measurement error, the bias that occurs when the used scale does not measure what it intended to measure. In addition, it is assumed that there is no high correlation between measurement errors and true scores, observed scores minus measurement errors, and errors scores are random and uncorrelated with each other. In practice, such assumption cannot be fully met. It is also assumed that items of the used scale must be essentially tau equivalent even though that perfect essentially tau-equivalence is seldom achieved (Cortina 1993).

Apart from essential tau-equivalence and uncorrelated errors, normality is also assumption of alpha (Sheng, Y., Sheng, Z. 2012). Moreover, Cronbach's alpha assumes unidimensionality, so when using multiple-dimensional scales, the researchers should compute Cronbach's alpha for each subscale as well as the entire scale. On this point, the present work stresses that when using scale that reflects different dimensions but not determined and defined yet, the researchers should first run a test such as a principle components analysis (PCA) to check the items and its own dimensions. However, such test does not work for pilot study purposes since factor analysis needs large sample size. From another standpoint, the present work recommends using Cronbach's alpha when the researcher composes a new scale and needs to check the overall reliability for a set of items, since such test describes the degree to which all the items of scale measure the same construct. However, when designing a new scale, the present work recommends that the researchers should first back to the related reviews of the intended scale to define the construct well. For understanding the topic of survey research design clearly and deeply, the present work invites the researchers to refer to the related sources (Griffee 2012).

Within this context, there are deep discussions with respect to the acceptable value of Cronbach's alpha amongst researchers. However, Griffee (2012) mentioned that a typical guideline is that 0.3 at the threshold, 0.5 or higher is adequate, and 0.7 or higher is high. George and Mallery (2003) provide more detailed categories of reliability values as rules of thumb (i.e., >0.9 "Excellent", >0.8 "Good", >0.7 "Acceptable", >0.6 "Questionable", >0.5 "Poor", while <0.5 "Unacceptable") (as cited in Khalid *et al.* 2012). From another standpoint, there are works in the literatures recommends using specific values of reliability according to the nature of the study since the reliability relies on a large extent on the use that is to be made of the results. For example, while Reid (1990) recommends that the reliability of

0.7 would be fair for survey instruments, some literatures recommend that the reliability should be about 0.9 or higher for important decisions (Cronbach 1990).

However, when using software (e.g., SPSS or SAS) for computing Cronbach's Alpha, the researchers had better check not only the overall value of alpha, but also the correlation between an item and other items of the used scale, since such way plays a key role in assessing how well item's score is internally consistent with composite scores from all other items, and then in deciding which item the researchers should revise or remove. De Vaus (2002) recommends that when the correlation is less than 0.3, then the correlation is weak and the researchers should remove that item. On this matter, the researchers can generally revise that item provided that re-pilot study again. A one thing the researcher generally should take into account that with increased sample size, the researchers can adopt value less than 0.3 (e.g., 0.2), since sample size affects the Cronbach's alpha. In this regard, the present work strongly emphasizes that the researchers should not remove all the offending items at once, but rather remove the one that will help them the most, then re-run the reliability, and repeat. In addition, the present work reminds the researcher to use the new values of the items, which already need reserve scoring. Otherwise, they will have negative values for those items.

On the other hand, some say that when the researchers use accepted scales, they should use them without any modifying even if those scales have problems with respect to new checking of the reliability. Their arguments behind that are, the researchers will able then to compare their results with the results of others who have used those scales, and they only improve the reliability when developing. From another standpoint, the present work tends to recommend modifying the accepted scales when having problems with respect to the reliability since the researchers may use those scales in different culture, different time or even there is problems in the Back-Translation step. Therefore, the present work strongly emphasizes, according the experience, doing the Back-Translation professionally. It, however, recommended previously what the researchers should do with respect to such issue.

2. Suggestions for further research: thoughts for professionalism

Even though reporting pilot studies is rare in the research literature (Teijlingen van *et al*. 2001), the study of Friedman (2013) provided reasons to report the results of pilot studies, for example but not limited to, the results of pilot studies might be useful to others; meta-analyses studies incorporate not just the data from the large studies, but from smaller ones including the pilot studies; finally, if the

pilot study does not lean to a full-scale trial, it may be particularly important for other researchers to understand why it did not. In fact, such reasons are logical. However, the present work strongly encourages and recommends that the researchers conduct and report their pilot studies fully since a well-conducted pilot study, in fact, can inform us about the strengths and weaknesses of the used scales as well as the initial outcomes.

Furthermore, since the sample size affect Cronbach's alpha, the researchers should consider that issue, especially, when conducting pilot studies. However, according to the experience, the present work strongly emphasizes following the next steps to report pilot study well especially in the research which uses questionnaire or survey as a main tool for collecting data. (1) Conduct a pilot study first taking into account the issues discussed above including checking the reliability by Cronbach's alpha, and put the initial results aside. (2) Conduct the main study taking into account checking the reliability again, and report the results. (3) Compare between those results (the initial and main ones), especially, with respect to the reliability. (4) Delete the items, which have common problems of reliability.

Reality, researchers face different research perspectives when conducting a research and some of them find such perspectives as obstacles, which make them confused to determine which way they should follow. The present work, however, finds such different perspectives as mercy, as it were, for researchers since they can chose, according to nature of the research and the condition of researchers themselves, what they find suitable for their research if they chose a clear and sound method.

3. Pilot studies in business sector: implementations and practices

The present research addressed in this work the importance of the pilot study for research and the most recommended practices of the most important quantitative methodological issues the researchers had better take into account when conducting the pilot study. However, the present work will also address in this section the importance of pilot studies for business sectors (e.g., business companies, banking sector, research centres and other) through addressing the most of its implementations and practices. On the other hand, talking about the implementations and practices of pilot studies in business sectors is really a wide topic since there are many different business sectors use such kind of studies. Therefore, the present research will address the most aspects of the implementations and practices of pilot studies in business sectors since such implementations and practices will consider as the common denominator amongst many different kinds of business sectors including marketing, and financial business sectors.

However, even though some consider doing pilot study is just a waste of time and money, others consider conducting pilot studies, especially, in business sectors is an important step since such kind of studies could help such sectors to achieve its goals not only effectively, but also efficiently (i.e., to reach goals with low costs). In fact, conducting a pilot study helps the business sectors to reduce or to avoid the number of unanticipated problems, which cost such sectors a lot, by overcoming difficulties and the risks that the pilot study reveals.

As mentioned above, there are indeed different business sectors which use pilot studies before embarking the main project or before taking the final decision about the intended project. For example, business companies, which need realistically analyze and studying the impact of expansion, developing existing products, or producing new products, generally use pilot studies for such purposes. Banking sector is also another example of business sectors which uses the pilot study to map out for the companies which need borrowing, for example. Other business sectors such as services and software business companies, for example, also use pilot studies to analyze and to study the new ideas and the new programs in the markets and to check then both of the pros and the cons of such ideas and products.

However, even though conducting a pilot study provides us with limited information comparison with the main study and does not guarantee success in the latter, but it almost always provide enough data for a business sector to decide at least whether to go ahead with the main project. Therefore, conducting a pilot study at first may help many business sectors to save a lot of time and money, especially, for those like market and medical research centres. Reality, For example, research centres (i.e., market and medical research centres) almost always rely on feasibility and pilot studies before embarking with the main project since such kinds of researches cost the business a lot in case embarking with the main project directly and without conducting the pilot studies first. So, due to the nature of such kinds of researches, such business centres rely on pilot and feasibility studies to check the markets first or to check the new product and then to give the green light to a business sector to go ahead with the main project to develop an existing product.

Moreover, conducting pilot studies helps a business sector to design the research protocol and to assess whether such protocol is realistic and workable. Such thing will help a business sector to determine then what resources of finance and staffs are needed for the main study. In addition, conducting a pilot study will play an important role to convince funding bodies that the intended research team is competent and knowledgeable, also to convince such bodies that the intended projects is feasible and worth funding. Then, to convince the stakeholders that the intended project

is worth supporting. On this point, Leon *et al.* (2011) mentioned that one of the most important implementations and practices in conducting a pilot study is that investigators, grant reviewers, and other stakeholders need to be aware of all sides of the pilot studies from to the essential elements to the limitations. Reality, the present work strongly tend to agree that such implementations and practices are reasonable and help business sector to reach goals with low costs in case the pilot or feasibility studies are done professionally.

All in all, different kinds of business sectors can use pilot studies to reduce the proportion of failed trials and allow funding bodies to finance the projects which feasibility has been demonstrated and quantified. Therefore, the present research strongly tends to agree what De Vaus (1993) recommended "Do not take the risk. Pilot test first".

References

Allison, P. D. 2001. *Missing data*. CA: Sage. 89 p.

Bajpai, S. R.; Bajpai, R. C. 2014. Goodness of measurement: reliability and validity, *International Journal of Medical Science Public Health* 3(1): 173–176.
http://dx.doi.org/10.5455/ijmsph.2013.191120133

Baker, T. L. 1994. *Doing social research.* 2nd ed. New York: McGraw-Hill Inc. 499 p.

Bennett, D. A. 2001. How can I deal with missing data in my study?, *Australian and New Zealand Journal of Public Health* 25(5): 464–469.
http://dx.doi.org/10.1111/j.1467-842X.2001.tb00294.x

Billingham, S.; Whitehead, A. L.; Julious, S. A. 2013. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database, *BMC Medical Research Methodology* 13(104): 1–6.
http://dx.doi.org/10.1186/1471-2288-13-104

Brislin, R. W. 1970. Back-translation for cross-cultural research, *Journal of Cross-Cultural Psychology* 1(3): 185–216.
http://dx.doi.org/10.1177/135910457000100301

Brown, J. D. 1997. Skewness and Kurtosis, *Shiken: JALT (Japan Association of Language Teaching) Testing* & *Evaluation SIG Newsletter* 1(1): 20–23.

Buhi, E. R.; Goodson, P.; Neilands, T. B. 2008. Out of sight, not out of mind: strategies for handling missing data, *American Journal of Health Behavior* 32(1): 83–92.
http://dx.doi.org/10.5993/AJHB.32.1.8

Bump, W. M. 1991. The normal curve takes many forms: a review of skewness and kurtosis, in *Annual Meeting of the Southwest Educational Research Association,* 24–26 January 1991, San Antonio, TX (ED 342790).

Cameron, A. C. 2004. *Kurtosis*, in M. Lewis-Beck, A. Bryman, T. F. Liao (Eds.). *The Sage encyclopedia of social science research methods*. CA: Sage. 545 p.
http://dx.doi.org/10.4135/9781412950589.n467

Cortina, J. M. 1993. What is coefficient alpha? An examination of theory and applications, *Journal of Applied Psychology* 78(1): 98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

Cramer, D.; Howitt, D. 2004. *The Sage dictionary of statistics: a practical resource for students in the social sciences*. CA: Sage. 188 p.

Cronbach, L. J. 1990. *Essentials of psychological testing.* 5th ed. NY: Harper and Row. 726 p.

De Vaus, D. 1993. *Surveys in social research.* 3rd ed. London: UCL Press. 379 p.

De Vaus, D. 2002. *Analyzing social sciences data: 50 key problems in data analysis*. London: Sage. 401 p.

Eekhout, I.; De Vet, H. C. W.; Twisk, J. W. R.; Brand, J. P. L.; De Boer, M. R.; Heymans, M. W. 2013. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level, *Journal of Clinical Epidemiology* (in press). http://dx.doi.org/10.1016/j.jclinepi.2013.09.009

Elliott, A. C.; Woodward, W. A. 2007. *Statistical analysis quick reference guidebook with SPSS examples.* 1st ed. London: Sage. 280 p.

Enders, C. K.; Bandalos, D. L. 2001. The relative performance of full information maximum likelihood estimation for missing data in structural equation models, *Structural Equation Modeling: A Multidisciplinary Journal* 8(3): 430–457. http://dx.doi.org/10.1207/S15328007SEM0803_5

Field, A. 2013. *Discovering statistics using IBM SPSS statistics.* 4th ed. CA: Sage. 952 p.

Field, A. P. 2005. *Discovering statistics using SPSS.* 2nd ed. London: Sage. 780 p.

Friedman, L. 2013. Commentary: why we should report results from clinical trial pilot studies, *Trials* 14(14): 1–2. http://dx.doi.org/10.1186/1745-6215-14-14

Garson, G. D. 2012. *Testing statistical assumptions*. NC: Statistical Publishing Associates. 51 p.

George, D.; Mallery, P. 2003. *SPSS for Windows step by step: a simple guide and reference*. *11.0 update*. 4th ed. Boston: Allyn & Bacon. 400 p.

Graham, J. W. 2009. Missing data analysis: making it work in the real world, *Annual Review of Psychology* 60: 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W.; Cumsille, P. E.; Elek-Fisk, E. 2003. Methods for handling missing data, in J. A. Schinka, W. F. Velicer (Eds.). *Research methods in psychology (2). Handbook of psychology*. NY: Willey. 118 p.

Griffee, D. T. 2012. *An introduction to second language research methods: design and data*. 1st ed. CA: TESL-EJ Publications. 213 p.

Heckman, J. J. 1979. Sample selection Bias as a specification error, *Econometrics* 47(1): 153–161. http://dx.doi.org/10.2307/1912352

Hoaglin, D. C.; Mosteller, F.; Tukey, J. W. 1983. *Understanding robust and exploratory data analysis*. NY: Wiley. 472 p.

Howell, D. C. 2007. *Statistical methods for psychology.* 6th ed. CA: Thomson Wadsworth. 739 p.

Kang, H. 2013. The prevention and handling of the missing data, *Korean Journal of Anesthesiology* 64(5): 402–406. http://dx.doi.org/10.4097/kjae.2013.64.5.402

Khalid, S.; Irshad, M. Z.; Mahmood, B. 2012. Job satisfaction among academic staff: a comparative analysis between public and private sector universities of Punjab, Pakistan, *International Journal of Business and Management* 7(1): 126–136.

Krithikadatta, J. 2014. Normal distribution, *Journal of Conservative Dentistry* 17(1): 96–97. http://dx.doi.org/10.4103/0972-0707.124171

Kuder, G. F.; Richardson, M. W. 1937. The theory of the estimation of test reliability, *Psychometrika* 2(3): 151–160. http://dx.doi.org/10.1007/BF02288391

Leon, A. C.; Davis, L. L.; Kraemer, H. C. 2011. The role and interpretation of pilot studies in clinical research, *Journal of Psychiatric Research* 45(5): 626–629. http://dx.doi.org/10.1016/j.jpsychires.2010.10.008

Olinsky, A.; Chen, S.; Harlow, L. 2003. The comparative efficacy of imputation methods for missing data in structural equation modelling, *European Journal of Operational Research* 151(1): 53–79. http://dx.doi.org/10.1016/S0377-2217(02)00578-7

Peng, C.-Y. J.; Harwell, M.; Liou, S. M.; Ehman, L. H. 2006. *Advances in missing data methods and imputations for educational research,* in S. Sawilowsky (Ed.). *Real data analysis*. Greenwich, CT: Information Age Publishing. 78 p.

Pett, M. A. 1997. *Nonparametric statistics for health care research: statistics for small samples and unusual distributions*. CA: Sage. 307 p.

Polit, D. F.; Beck, C. T.; Hungler, B. P. 2001. *Essentials of nursing research: methods, appraisal, and utilization.* 5th ed. Philadelphia: Lippincott Williams and Wilkins. 544 p.

Porta, M. 2008. *A dictionary of epidemiology.* 5th ed. Oxford: Oxford University Press. 320 p.

Razali, N.; Wah, Y. B. 2011. Power comparisons of Shapiro-Wilk, Kolmogrov-Smirnov, Lilliefors and Anderson-Darling tests, *Journal of Statistical Modeling and Analysis* 2(1): 21–33.

Reid, J. 1990. The dirty laundry of ESL survey research, *TESOL Quarterly* 24(2): 323–338. http://dx.doi.org/10.2307/3586913

Rubin, L. H.; Witkiewitz, K. S.; Andre, J.; Reilly, S. 2007. Methods for handling missing data in the behavioral neurosciences, *Journal of Undergraduate Neuroscience Education* 5(2): 71–77.

Schafer, J. L. 1999. Multiple imputation: a primer, *Statistical Methods in Medical Research* 8(1): 3–15. http://dx.doi.org/10.1191/096228099671525676

Schafer, J. L.; Graham, J. W. 2002. Missing data: our view of the state of the art, *Psychological Methods* 7(2): 147–177. http://dx.doi.org/10.1037/1082-989X.7.2.147

Schlomer, G. L.; Bauman, S.; Card, N. A. 2010. Best practices for missing data management in counseling psychology, *Journal of Counseling Psychology* 57(1): 1–10. http://dx.doi.org/10.1037/a0018082

Sekaran, U.; Bougie, R. 2010. *Research methods for business: a skill building approach.* 5th ed. New Delhi: John Wiley. 488 p.

Sheng, Y.; Sheng, Z. 2012. Is coefficient alpha robust to nonnormal data?, *Frontiers in Psychology* 3: 1–13. http://dx.doi.org/10.3389/fpsyg.2012.00034

Tabachnick, B. G.; Fidell, L. S. 2007. *Using multivariate statistics.* 5th ed. Bosten: Allyn and Bacon. 1008 p.

Tavakol, M.; Dennick, R. 2011. Making sense of Cronbach's alpha, *International Journal of Medical Education* 2: 53–55. http://dx.doi.org/10.5116/ijme.4dfb.8dfd

Teijlingen van, E.; Rennie, A. M.; Hundley, V.; Graham, W. 2001. The importance of conducting and reporting pilot studies: the

example of the Scottish Births Survey, *Journal of Advanced Nursing* 34(3): 289–295.
http://dx.doi.org/10.1046/j.1365-2648.2001.01757.x

Warner, R. M. 2008. *Applied statistics: from bivariate through multivariate techniques*. CA: Sage. 1101 p.

West, S. G.; Finch, J. F.; Curran, P. J. 1996. Structural equation models with nonnormal variables: problems and remedies, in R. H. Hoyle (Ed.). *Structural equation modeling: concepts, issues and applications*. CA: Sage. 75 p.

Osama A. HAZZI. A PhD candidate in Business Administration (Organizational Behavioral Studies), Faculty of Economics, Damascus University. The author of three research papers. Reviewer at two peer-reviewed international journals. Research interests: Human Resource Management, Organizational Citizenship Behavior, Organization Justice.

Issa Sh. MALDAON. Associate Professor at Department of Business Administration, Faculty of Economics, Damascus University, Syria. Continuous Training manager at the National Institute of Public Administration (Institut National d'Administration "INA", Syria). The author of many research papers, books, and presented papers in regional and local conferences. Deputy Minister of Social Affairs and Labor (previously), Syria. Deputy Dean for Scientific Affairs at Faculty of Economics (previously), Damascus University. Deputy Dean for Administrative Affairs at INA (previously), Syria. Research interests: Human Resource Management, Strategic Management, Leadership.