INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2222-6990

Testing on the Validity and Reliability of Task Based Language Teaching Questionnaire

Chow Kuok Feng & Hamidah Yamat

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v9-i12/6743 DOI: 10.6007/IJARBSS/v9-i12/6743

Received: *10 November 2019,* ***Revised:*** *27 November 2019,* ***Accepted:*** *12 December 2019*

Published Online: 29 December 2019

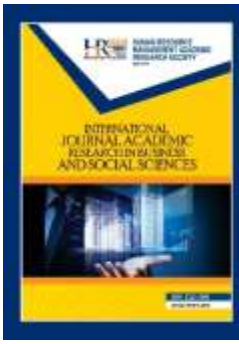
In-Text Citation: (Feng & Yamat, 2019)

To Cite this Article: Feng, C. K., & Yamat, H. (2019). Testing on the Validity and Reliability of Task Based Language Teaching Questionnaire. *International Journal of Academic Research in Business and Social Sciences*, *9*(12), 474–485.

Vol. 9, No. 12, 2019, Pg. 474 - 485

http://hrmars.com/index.php/pages/detail/IJARBSS | **JOURNAL HOMEPAGE**

Full Terms & Conditions of access and use can be found at
http://hrmars.com/index.php/pages/detail/publication-ethics


INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2222-6990

Testing on the Validity and Reliability of Task Based Language Teaching Questionnaire

Chow Kuok Feng & Hamidah Yamat
Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Abstract
Task Based Language Teaching (TBLT) has been widely used in the context of second language acquisition which promotes actual language use. The purpose of the postgraduate research study is to identify primary school English teachers' perception on TBLT. Thus, in order to collect the significant data, a survey questionnaire is carried out to gather the useful information. Therefore, this pilot study is conducted before the final study intended to explore any methodological issues arising. Hence, a survey questionnaire was developed and the purpose of this study is to test the 'Primary School English Teachers' Perception on TBLT' survey questionnaire. The findings of the study focused on face validity, content validity, test-retest validity and internal consistency reliability. Some positive and negative responses were received throughout the study. In a nutshell, the instrument has a high face and content validity, fairly high correlation coefficient as well as high level of Cronbach's alpha.
Keywords: Task Based Language Teaching (TBLT), Validity, Reliability, Questionnaire, Validity and Reliability Test

Introduction

Carrying out a research is definitely not an easy task and surely there will always be problems surfacing one after another throughout the entire process, especially when it comes to data collection part. One of the most common issues that researcher would probably face during the data collection phase is the methodological problem. Matters such as absence of validity or reliability of the instrument is always seen as an obstacle. However, most of the issues arise during the process of data collecting could be clearly avoided if a pilot test was first carried out to run through and to check on the instruments beforehand. Therefore, in order to avoid any unforeseen circumstances or flaws which might arise during the process of data collection, it is crucial for researcher to actually run a pilot test on the instruments used. Hence, the need to conduct this pilot study is to determine that issues like methodological flaws could be avoided.

The purpose of this pilot study is to examine the survey questionnaire which will be applied in the actual data collection in the final full-scale study. Nashwa (2018) suggests that

when piloting a study, a researcher must pay attention to a few important aspects so that the instrument used is applicable to obtain significant data. Among them are to determine what methods used are best for achieving the research questions, estimate the duration of time and what resources will be needed to complete the larger final version of the study. Hence, for this pilot study, the researcher planned to test the 'Primary School English Teachers' Perception on TBLT' survey questionnaire which was adapted from Joen and Hahn's Teacher Questionnaire (2006). The finding of this study could definitely help the researcher to modify and further adapt on the methodology aspect before it was used in the actual study.

The structure of this study begins with revealing the purpose of instruments, then explaining on how the instrument was developed. Next, followed by piloting the instrument, then it proceeds with the validity and reliability test. Lastly, the study exposes the finding of study.

This pilot study was carried out not just to simply check on the perception questionnaire in order to avoid methodological issues, however, it was to also contribute other researchers with better understanding on the importance of validity and reliability of an instrument as well as its methodological features. Furthermore, this pilot study can also be used as guidance and demonstration for those researchers who wish to implement pilot test on their instrument such as questionnaire.

Literature Review

Japheth (2014) reckons that measurement of instrument refers to different means through which a researcher gets data from the participants for his or her research work. Basically, instrument acts as a tool used to collect, measure and also analyses data related to the subject. There are plenty versions of instruments that can be used in the study and it depends on the nature of research that is going to be carried out. Among them are surveys, observations, field note, interviews and etc.

For the actual research, the researcher applies survey research in order to obtain the desire result. Creswell (2014) states that survey research designs is a quantitative research in which the researcher run a survey such as questionnaires to a sample or to the entire populations of people to describe attitudes, opinions, behaviors and etc. Therefore, based on the objectives of the study mentioned earlier, the researcher is trying to identify the perceptions of TBLT among primary school English teachers such as their understandings of TBLT concepts, views on TBLT implementation, and reasons for them to choose, or avoid implementing TBLT in classroom. Hence the survey research design seems to be appropriate to be applied in this research.

A questionnaire is a research instrument whereby a set of questions are listed for the purpose of gathering useful and significant information from the respondents. The questionnaire is the favored tool in research as it can often provide a cheap yet effective way of collecting data in a structured and manageable form. Syed (2016) indicates that to collect data through questionnaire is relatively cheap and economical since the researcher does not require to spend so much to during the preparation. Despite of being affordable and easy to administer, questionnaire can also cover many subjects or issues or be very simple and focus on one important area. It is a convenient way of collecting useful comparable data from a large number of individuals.

A perception questionnaire adapted from Joen and Hahn's Teacher Questionnaire (2006) is used to measure teachers' perception of TBLT among primary school English teachers. The researcher was very detailed in developing the instruments such as creating the title, constructing proper introduction and making meaningful questions so that it could produce valid and significant result. The survey questionnaire was constructed precisely and the researcher had also taken different important aspects into consideration.

The Title

All the questionnaire requires a title and it is advisable that the title should be short, simple, appealing as well as inviting. It does play a vital role to help readers in creating the initial impression, hence to decide if there is a need to further look into the detail of the questionnaire. Title used must not only be appropriate with the items asked, it must also be able to catch and hold the participants' attention and attempt the items accurately. Sandeep (2016) posits that researchers should take efforts to choose an informative and appropriate title when creating questionnaire.

Instructions

Instruction on the questionnaire is very crucial as it tells the respondents what to do or rather how to respond to the items. Since a postal questionnaire for self-completion was administered, hence a clear and accurate instruction was used so that the respondent will know how to react to the items.

Likert-type Items

Likert-type is one of the most common style used in questionnaire. This type of instrument requires the respondent to give the respond according to a pre-defined scale, ranging from strongly agree (SA), agree (A), disagree (D) and strongly disagree (SD). It is designed to offer a real or hypothetical situation based on the research study. Basically, participants are required to demonstrate their level of agreement with the given statements on a metric scale. Ankur, Saket, Satish, & Pal (2015) suggest that Likert type items is rooted into the aim of the research and sometimes the purpose of the research is to understand about the opinion and perceptions of participants related with single 'latent' variable. Hence, the items asked through Likert-type style must be closely related to the research questions in order to get the most appropriate respond from the participants.

Multiple Choice Question

Whereas for multiple choice question, there are a few advantages that researcher can be benefited from it. First, multiple choice question allow researcher to have a quick and easy access to the desirable result from the participants. Most of the questionnaires include questions which provide a number of predefined responses as it permits the researcher to hold some control over the responses given (David & Peter, 2003). Through multiple choice question, participants are able to select the options which best describe their thoughts. By doing so, it helps the researcher to seize the chance to explore the research questions further. Besides, multiple choice questions

style appears to be more user friendly. This type questionnaire style usually comes in simple and short manner as the participants tend to ignore or skip the items if it is to be too complex or lengthy, respondents might just ignore or skip the items. However, items constructed through multiple choice question should not only be brief but also able to cover research questions which will provide significant insight for the researcher. Weimer (2018) adds that multiple choice questions must allow the researcher to covers lots of content areas on a questionnaire and still be answered in a short amount of time.

Research Question

To test the validity and reliability of English teachers' perception on Task Based Language Teaching questionnaire.

Research Objectives

1. What is the validity of English teachers' perception on Task Based Language Teaching questionnaire?
2. What is the reliability of English teachers' perception on Task Based Language Teaching Questionnaire?

Procedures

Before a researcher can disseminate the questionnaire, it is important to pilot the draft questionnaire. It is also wise to never implement a questionnaire which has not been conducted or piloted. Technically, a pilot stage will allow the researcher to ensure that all the appropriate issues are included, the order is well organized, level of ambiguity or leading questions are identified, the pre-codes are correct and etc.

When it comes to measurement, it is advisable to involve assigning scores to individuals so that they represent some characteristic of the individuals. After developed and constructed the survey questionnaire, the researcher used Google Form to carry out the survey. To pilot the questionnaire, the researcher first posted a link to the Google Form to the participants. Since the purpose of this study is to pilot the instrument, hence the researcher disseminated the questionnaire to only 10 participants from 2 different schools. The participants chosen were different from the main study but they were from the similar setting. The reason why the researcher decided to select different participants is because the same participants might not be interested in the exact same questionnaire later on when the questionnaires are administered for the actual study. Therefore, the findings might not be valid. Janghorban et al. (2014) states that asking the exact same person to answer same question multiple times may cause them to lose interest in the study which is known as semantic satiation hence, the researcher tried to avoid this from happening in the further actual research by choosing only 10 different participants.

A period of time was given to the participants to complete the questionnaire. After the researcher had obtained the result, he started to code and analysis the data. Most questionnaire required respondents to respond to questions by ticking in boxes provided. The researcher then coded the answers. For the Likert-type items, each numerical score was coded for every item.

For instance, strongly agree (SA)=5, agree(A)=4, neutral(N)=3, disagree(D)=2 and strongly disagree (SD)=1. Whereas for open-ended items, a value of '1' was given for every selected item and a value of '2' was awarded for every unselected reason.

Statistical Package for Social Sciences (SPSS) and Microsoft Excel were used to analyse the data gathered through questionnaires and the data was also tabulated into a table to ease its presentation.

Methodology

Research Design

Validity

Ghauri and Gronaug (2005) explains that validity as how well the gathered data covers the actual range of study. In short, validity can be assumed as the extent to which the scores from a measure represent the variable they ought to. Hence, the instrument used could be reliable and it should also be valid at the same time. Sometimes, a measure can be very reliable but have no validity or whatsoever. Therefore, a drafted questionnaire should always be ready for establishing validity.

For the pilot study, the researcher had divided validity into several different types. Among them are face validity and content validity.

Face Validity

Face validity is a subjective judgment on the operationalization of a construct. (Hamed, 2016). Face validity is the degree to which a measure appears to be related to a specific construct. Therefore, it could be explained as the degree that the respondent judges that the items of the instrument are suitable to the targeted construct and the research objectives. The judgment of the instrument could be of experts like other experience researcher or test takers. Basically, face validity helps to evaluate the appearance and the layout of an instrument in term of its feasibility, readability, formatting and last but not least, the clarity of the words used.

In short, Oluwatayo (2012) implies that face validity refers to the relevance of the measuring instrument as to whether the items of the instrument seem to be relevant, the items constructed are rational, unambiguous and clear. Hence, for examining the face validity purposes, the researcher first presented the questionnaire to an expert, whom happened to be his supervisor of the research study. The researcher shown and discussed the questionnaires in terms of clarity of wording, layout and style as well as likelihood the respondent would be able to answer the questions.

Content Validity

Content validity is the extent to which measure covers the construct of interest. Yassir et al. (2016) suggests that content validity can be assumed as an evaluation of each test item in order to make sure it is relevant to the intended construct, if the items are well or correctly worded or whether the scoring or scaling is appropriate to ensure the items on the instruments are representative samples. Hence, it is crucial to predict the efficacy of the tool in order to lower the measurement error that could happen during administration.

It is debatable in some testing when the items in the questionnaires do not seem to measure what they are intended to measure, therefore, Brennan & Hays (1992) state that in order to avoid vague and infeasibility of an instrument, an instrument must undergo its content validity. Therefore, in order to check on the instrument content validity, the researcher must first establish a qualitative review of the instrument by some expertise to gauge the feedback based on 5 questions:

1) Do you have any suggestions regarding the definition of the constructs?
2) Do the test items appear to cover the full topic of content within each construct?
3) Are the items in the instrument clearly phrased and not ambiguous?
4) Are the items constructed in the instrument appropriate for teachers?
5) Do you have any new items that you would wish to include?

Reliability

Reliability refers to the degree of consistency of a measure. A test will be reliable when it gives the same repeated result under the same condition. Mujis (2011) defines reliability as 'the extent to which test scores are free from measurement error'. According to Creswell (2014), she posits that there are many kinds of reliability depending on the amount of times the instruments are administered and the amount of individuals who deliver the information. For this study, the researched looked into 3 types of reliability, which are test-retest reliability, internal consistency.

Test-retest Reliability

Test-retest reliability is crucial before administering the instrument as it allows the researcher to know if it is able to yield the similar results across time. When researchers measure an instrument that they assume to be consistent, hence the scores obtained should also be consistent across time. Huck (2007) imposes that checking on reliability is crucial because it refers to the consistency across the parts of measuring insturment as it indicates the repeatability of test scores within a period of time.

In order to measures the stability of scores across time, the researcher first administered the instrument on the 10 participants which were chosen early, then the researcher collected the data obtained. After two weeks, the researcher administered the exact same instrument on the same group of participants again and the data was gathered.

Internal Consistency Reliability

Mujis (2011) indicates that the internal consistency reliability is looking at the correlation between all items that make up the constructs to ensure that the items are measuring the same concepts. The researcher uses this test when he or she wants to know if all the items in the questionnaires are related and reflect the concept. The most commonly used internal consistency measure is the Cronbach Alpha coefficient.

In order to examine the internal consistency reliability, the researcher decided to use Cronbach Alpha coefficient. It is viewed as one of the most appropriate measure of reliability when making use of Likert scales (Whitley 2002). Therefore, the researcher administered the instruments and the data obtained was organized and analysed using SPSS.

Findings and Discussion

Face Validity

Based on the discussion, some reviews and comments were given in order to improve the validity of the instrument. Firstly, the layout and the design of the questionnaire is pleasantly presented with clear and brief introduction. However, there were some adjustment made in order to improvise the questionnaires. The general demographic information was removed as it did not serve any purpose in the findings of data. The justification of removing that information is because they are less likely to provide any useful or rather desire data to the research. It is advisable to only keep those useful and meaningful items throughout the questionnaire so it would be easier for researcher to analyse the data later on.

Content Validity

The researcher listed the question and presented the questionnaire to experts. Upon receiving the responses and comments, the instrument was immediately reviewed and revised accordingly. Appropriate modification was applied to the questionnaire. The questionnaire was divided into three different section, namely Section A, B and C. For Section A and B, the comments received were relatively positive and no changes were made. However, for Section C, a little amendment was made as there was some ambiguity in the questionnaire item. For this particular section in the survey questionnaire, item no. 25, 'I have limited target language proficiency and limited knowledge of task-based instruction.' were found to be ambiguous. Limited target language proficiency and limited knowledge of task based instruction were two different aspects, so it would be contradicting one another if they were to be constructed in a sentence. Therefore, it should not be combined into only one sentence.

The researcher divided the sentence into two separated items which are 25) 'I have limited target language proficiency.' and 26) 'I have limited knowledge of task-based instruction.' The justification for the changes is because one could have sufficient knowledge of task-based instruction but limited target language proficiency and vice versa. So the respondents might be confused by the item itself if the respective item was not separated to form 2 different items.

Test-retest Reliability

The data obtained was tabulated and analysed using Microsoft Excel 2010 to find the reliability coefficient between the two scores. Then, the researcher looked at test-retest correlation between the two sets of scores obtained.

Figure 1.0: Value of test-retest correlation between first administered questionnaire and second administered questionnaire

Subject	First Administered	Second Administered
1	55	54
2	76	74
3	75	72
4	65	65
5	52	55
6	70	70
7	74	74
8	70	69
9	54	54
10	73	72
Corr Coef	**0.989808575**	

The researcher looked at the reliability of a test based on the reliability coefficient. Letter 'r' was denoted and it was conveyed into a number ranging from 0 and 1.00, with r=0 indicates no reliability and r=1.00 means excellent reliability. Therefore, it could be concluded that the larger the reliability coefficient, the more reliable the instrument will be. Figure 1.0 shows the value of test-retest correlation between the questionnaires on primary school teachers' perception on TBLT which was administered for the first and second time, two weeks apart. The result revealed that r is =.98 which was fairly high. The results yielded significant score after the instrument was piloted for its reliability. Hence, it is safe to say that the instrument is reliable as the higher the correlation value gives more reliable to the instrument.

Internal Consistency Reliability

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.892	.856	15

Figure 2.0: Value of Cronbach's Alpha of the instrument

A common accepted value of an alpha is 0.7 or 0.6. However, value of 0.8 or higher is also consider as good reliability. Based on the reliability statistic shown from the above, the value of Cronbach's alpha is .89 in which a fairly high level of internal consistency reliability was presented within the piloted instrument. Therefore, the internal reliability of the instruments adapted is seemingly high and could be administered for the actual study.

Conclusion

Generally, one of the most substantial major findings discovered from the pilot study was the significant result of the validity as well as the reliability of the instrument which used to gauge the perception of English teacher in using TBLT in their teaching. Despite of the adaptation of the perception questionnaire through Joen and Hahn's Teacher Questionnaire (2006), the researcher planned to pilot the instrument in order to further discover its methodological aspects.

In a nutshell, the pilot study shown that the validity test such as face and content validity has indicated that the instrument is appropriate for administering in the actual study and some changes were made to improve the validity of the test.

Besides that, the instrument also went through test-retest reliability test to check on its correlation coefficient. r=.98 shown that the instrument has a high correlation coefficient. Cronbach's Alpha was also used to test on its internal consistency reliability and α=.89 indicates that the instrument has a high reliability.

Designing a questionnaire for a researcher is definitely not an easy task as it involves a lot of important aspects that the he or she needs to take into consideration. In fact, sometimes it is easy for the researcher to overlook errors and ambiguous question layout and construction. Therefore, piloting the instrument becomes one of the vital steps for the researcher to determine if the data collected helps the researcher to meet the objectives of the study as well as test the validity and reliability of the instrument.

During the piloting stage, mistakes can be spotted quickly and rectified relatively easily. Besides that, ambiguous questions can also be restated or redeveloped once the instrument has been piloted. Luiz (2017) states that a pilot study helps to provide information like weather or not a full-scale study will be feasible and state all the possible recommendation and amendments for the research instruments.

Suggestions and Recommendations

Although the study displayed a high level of validity and reliability of the instrument, some useful suggestions and recommendations could help to escalate and enhance the quality of the instrument. Perhaps one of the beneficial suggestions to further increase the attribute of the perception is to include or add in more meaningful questions as part of the items in the instruments.

In fact, items used in the perception questionnaire should always allow the participants to have a better understanding on the entire questions asked in order for them to provide their significant insight as well as opinions. Therefore, in order to obtain more representative data, increasing the items in the questionnaire appropriately could possibly optimize the chances of the researcher to acquire more meaningful view, hence strengthen the outcome of the research.

Besides that, based on the findings of the study, it is also important that the researcher must try to avoid ambiguous and confusing questions in their questionnaire. Therefore, it is crucial that they should always be alert and wise when it comes to items settings. It is recommended that the researcher need to constantly equip themselves with sufficient knowledge regarding to the topic that they are researching on. One of the simple ways is to look for materials such as articles or journal which is closely related to the researcher's topic.

Acknowledgement
The researcher would like to thank Dr Hamidah Yamat for her dedicated guidance and assistance. The research also wishes to express his genuine gratitude to Dr Hamidah Yamat for sharing valuable knowledge throughout the course of this research.

Corresponding Author
Dr Hamidah Yamat, Associate Professor of TESL in National University Malaysia.
Email: hamidah_yamat@ukm.edu.my

References

Ankur, J., Saket, K., Satish, C., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 397-403

Brennan, P. H., & Hays, B. J. (1992). Focus on psychometrics the kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. *Research in Nursing & Health*, *15*(2), 153-158.

Creswell, J. W. (2014). *Eudcational Research: Planing, Conducting and Evaluating Quantitative and Qualitative Research Fourth Edition*. Great Britian: Pearson.

David, W., Peter, B. (2003). *Using research instruments: A guide for researchers*. New York, NY: Taylor & Francis Group

Hamed, T. (2016). Validity and reliability of the research instrument; How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management (IJARM)*, *5*(3), 25-36.

Huck, S. W. (2007). *Reading Statistics and Research*, United States of America: Allyn & Bacon.

Janghorban, R., Latifnejad, R., & Taghipour, A. (2014). Pilot Study in Qualitative Research: The Roles and Values. *Journal of Hayat*: *19*(4).

Japheth, A, Y. (2014). *Choosing the right measurement instrument for your project: Tip to apply*. Retrieved from https://nairaproject.com/blog/measurement-instrument.html [30 May 2019]

Luiz, C. (2017). *What is a pilot study?* Retrieved from *of Educational and Social* Retrieved from https://www.students4bestevidence.net/pilot-studies/ [25 May 2019]

Mujis, D. (2011). *Doing Quantitative Research in Education 6th Edition*. London: SAGE Publications Ltd.

Nashwa, I. (2017). Pilot study, does it really matter? Learning lessons from conducting a pilot study for a qualitative PhD thesis. *International Journal of Social Science Research*, *6*(1), 1-17

Oluwatayo, J. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research, 2*, 391-400.

Sandeep, B. (2016). Formulating the right title for a research article. *Journal of The Association of Physicians of Indian*, 64, 53-55.

Syed, M. S. K. (2016). *Basic guidelines for research: An introductory approach for*

all disciplines. Bangladesh: Book Zone Publication.

Yassir, Y. A., McIntyre, G. T., & Bearn, D. R. (2016). Three questionnaires to assess the perception of fixed orthodontic therapy before, during and after treatment: Validity and reliability. *European Journal of Orthodontics*, *39*(4), 402–410. doi: 10.1093/ejo/cjw076

Weimer, M. (2018). *Advantages and disadvantages of different types of test questions.* Retrieved from https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/

Whitley, B. E. (2002). *Principals of research and behavioural science*. Boston, McGraw-Hill.