

# Evaluation of a Data Collection Instrument for “Students Learning Styles in an EFL Context”: A Case Study

Mafuta Mungombo Claude<sup>1</sup>

<sup>1</sup>University of Kinshasa/ Unikin

Publication Date: 2025/05/12

**Abstract:** This article focuses on the assessment of data collection instruments designed to collect data on the topic “STUDENTS LEARNING STYLES IN AN EFL CONTEXT: A CASE STUDY ” for its validity and reliability. The evaluation of the instrument was done by 4 teachers whose 3 males and 1 female of the Languages School of the Faculty of Letters and Human Sciences who were supposed to be knowledgeable on the topic. The instrument used the mixed methods i.e. quantitative and qualitative methods. The qualitative method was converted into quantitative to obtain a more straightforward determination of the overall quality of the instrument. The study shows that there were some questions to improve and to be reformulated. Overall, from what our respondents have said, we can say that the instrument was valid and be administered to respondents for data collection on this issue.

## ➤ Resume:

*Cet article se concentre sur l'évaluation de l'instrument de la collecte de données conçu pour collecter les données sur le sujet “LE STYLES D'APPRENTISSAGE DANS UN CONTEXTE D'ANGLAIS COMME LANGUE ETRANGERE : CAS D'ETUDE ” pour sa validité et fiabilité. L'évaluation de l'instrument a été faite par 4 Enseignants dont 3 hommes et 1 femme de l'Ecole de Langues de la Faculté des Lettres et Sciences humaines qui sont supposés avoir des connaissances sur le sujet. L'instrument a utilisé les méthodes mixtes c-à-d. la méthode quantitative et la méthode qualitative. La méthode qualitative a été convertie en méthode quantitative pour obtenir une détermination la plus franche de la qualité globale de l'instrument. Après l'évaluation de l'instrument par les répondants, il ressort qu'il y a quelques questions à améliorer et à reformuler. De manière générale, de ce que nos répondants ont dit, nous pouvons dire que l'instrument était valide et peut être administrés aux répondants pour la collecte de données sur cette question.*

**Keywords:** Learning Style, EFL.

**How to Cite:** Mafuta Mungombo Claude. (2025). Evaluation of a Data Collection Instrument for “Students Learning Styles in an EFL Context”: A Case Study. *International Journal of Innovative Science and Research Technology*, 10(4), 3277-3284. <https://doi.org/10.38124/ijisrt/25apr1415>.

## I. INTRODUCTION

The current article focuses on assessment of the data collection instruments designed to collect data on the topic “STUDENTS LEARNING STYLES IN AN EFL CONTEXT: A CASE STUDY

Depending on the nature of the information to be gathered, different instruments evaluation could be used to conduct such an assessment: evaluation forms for gathering data from official 1 sources surveys/interviews to gather information from target, population and focus groups to obtain qualitative evaluations.

As far as this assessment was concerned, we used self-administered item evaluation form to capture information from a small population of evaluators. In one sense, this process is flexible, making it possible to present the items of

the research instrument on the right side of the form and check of the items for quality on a scale on the left side of the form. On the other hand, scaled items often represent the most common choice in assessing people's attitudes, orientations, experiences and perception about a phenomenon.

This process was our best choice as long as it allowed us the possibility of relating the perceptions (judgments) of each evaluator to the perceptions of all others.

## II. RELATED LITERATURE

Decisions as to which statistical analysis to conduct has been an issue in studies of raters agreement or reliability. Suggestions have been that the traditional correlations, such as the Pearson correlation or the simple Spearman's Rho or ranked data correlations are inaccurate measures and many researchers suggest considering the intra-class correlation

coefficient (ICC), all of which can be calculated using the SPSS scale reliability procedure. A debate continues as to which measures is the most appropriate though McCray (McCray, 2013) recommends the Gwet's statistic (Gwet, L. , 2010; 2008) as being preferable to the Cohen's Kappa statistics (Cohen, 1960) when nominal statistics are concerned.

Interrater agreement or reliability has been of great concern in a variety of fields particularly where performance rating by raters has been the preferred practice in evaluation. Such are the areas of curative as well as preventive of rehabilitative medicine and nursing (Havercamp, Kimberlin and Winterstein, 2008; Hallgren , 2012) and education (Graham et al. 2012) to name only a few.

This research aims at assessing the validity and enhancing reliability of the instruments designed to collect data on the research topic **“STUDENTS LEARNING STYLES IN AN EFL CONTEXT: A CASE STUDY ” IN AN EFL CONTEXT: CASE OF FIRST YEAR AIA**. The instrument assessment involved four teachers from Languages School of the Faculty of Arts of the University of Kinshasa who were supposed to be knowledgeable on the topic.

#### ➤ Hypothesis

This research raises one hypothesis: the instruments used to collect data on the topic above are not reliable. Given the methodology used in this research, the hypotheses to test are that there is no relationship between the raters' scoring of the items, the quality of (form/content) was (not) good, there are (no) suggestions for the improvement of the items.

### III. METHODOLOGY

#### A. Research Design

Until recently, a great majority of research studies were based either on quantitative or qualitative research methods (Bryman, 2006: 97-113). In the course of time, there emerged a group of “purists” on each side (Denzin & Lincoln, 2000: 1-32; Johnson & Onwuegbuzie, 2004: 14-26) advocating for either one paradigm of the quantitative research have been maintaining that the research designs should be organized on objective and concrete plans. On the other hand, the proponents of qualitative research claimed that the research should be based on interpretive and “hermeneutic” approaches (Johnson & Onwuegbuzie, 2004: 112-133).

However, it is clearly known that qualitative and quantitative research methodologies have both inherent strengths and weaknesses. For example, quantitative research is very useful for making generalizations about populations, but it cannot profoundly address respondents' internal perspectives. Conversely, qualitative research supply researchers with “in-depth and rich information about participants viewpoints” (Johnson & Christensen, 2004:410), but its findings cannot be easily generalized due to small size of the samples. Thus, it is now clearly known that both quantitative and qualitative research methods have some natural shortcomings, and these shortcomings have led to the birth of a new third research paradigm called “mixed methods

research, mixed research, mixed methodology, and multimethod research”. Johnson, Onwuegbuzie and Turner (2007:120) define mixed methods research as “a class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study or set of related studies”.

The power of mixed research as a third research paradigm springs from the fact that it creates a “practical synthesis” for the researcher to utilize both quantitative and qualitative research methods, in this way, the researcher can ensure “complementary strengths and non overlapping weaknesses” (Johnson & Onwuegbuzie, 2007:3).

In this research, the quantitative (evaluation) research method was used for making generalizations of scores that are evaluative, i. e. qualitative. But the evaluative statements by raters were converted into scaled numerical data to obtain a more straightforward determination of the overall quality of the instrument.

#### B. Participants

The participants in this research who are the raters were teachers at the Languages School of the Faculty of Arts of the University of Kinshasa who are considered as experts in teaching English language with new technology. As all the teachers of English use new technology in their teaching, the teachers understood the situation of the instruments application to qualify as the expert judges of the instruments quality. In terms of gender, the raters included both males and females.

#### C. Data Collection Instrument

##### ➤ The Items Rating Form

The form presented 14 items to be rated on a scale from poor to excellent in the likert scale format. The scoring terms were arranged from the most negative or lowest (poor) to most positive judgment or highest (excellent). Thus, the rubrics were converted the numerical values 1 (very poor), 2(poor), 3(quite good), 4(very good) and 5(excellent).

Although all items were presented in one single rating form, the instrument items to be rated had been presented in 3 sections: the identification of the participants (item1, item2 and item 3), purpose of use of video episodes, audios and textbook (item1, item 2, item3, item6, item7, item9, and item11) and the attitude towards the use of video episodes, audios and text book (item 4, item 5, item 8, item 10).

The second part of the questionnaire, video episode, audios and textbook use and Literacy, was developed by the researcher in order to find out for what purposes English language teachers and students used video episode, audios and textbook, to what extent they were to maximize their language teaching and learning practices with technological tools, and how much they were aware of the contemporary technologies in the context of language teaching and learning.

There was five major items in the second section of the questionnaire. The first item tried to explore for what purposes English language teachers and students use video episodes, audios and textbook. The second item explores how many hours the participants used video Episodes, audios and textbook. The third and forth items was open-ended questions related to the use of English video episodes, audios and textbook.

The third section of the questionnaire, Attitudes toward the use of English video episodes, audios and textbook were used in order to explore the perceptions of the respondents about the use of English video episodes, audios and textbook and their attitudes toward computers in general.

There was a total of 5 items in this part of the questionnaire which explored the attitudes of the participants towards the use of English video episodes, audios and textbook. Through this scale, the opinions of the teachers and students about the use of English video episodes, audios and textbook and roles of computers in education were collected.

#### D. Data Analysis:

The validity of the items used in this research was established through the mean score and the correlation of the scores to the instrument items by the four instrument evaluators. An inter-rater reliability for multiple raters (4 in this case) was required for that purpose. Both the usual Pearson and Spearman Rho correlations were run to identify any relationship between the scores. However, since the Spearman or the Pearson Rho are considered as quite inaccurate measures of consistency, the Reliability

coefficient (Cronbach's alpha) was used, paired with the intraclass coefficient as suggested by the SPSS procedure. The means for the scores were provided as additional evidence of the quality of the evaluated items. The Spearman and the Rho were calculated for comparison as part of the pedagogical exercise the researchers were involved in. Tables of the calculated coefficients in the different procedures were drawn to ease the comparison.

#### E. Indings

Beside the sample characteristics of gender, categorization was obtained from the obtained responses.

##### ➤ Sample Characteristics

The 4 teachers who assessed the instrument included both males and females. There were 3 males and 1female from the Language School mentioned above. Almost all the raters were married and one was single. All of the raters were Congolese.

##### ➤ Mean Scores for the Items

As far as validity of the instrument is concerned, there was a total of 15 items, out of which 7 items were rated excellent or items 3, 4, 5, 6, 11, 12 and 13; five items were rated very good or items1, 7,9,10 and 14; two items were rated good or items 8 and 15; one item was rated poor or item 2. Comparatively to what is said above, we can say that the majority of the experts proved positively the instrument.

The scores were averaged for each item and a table of means was made as indicated in the table below (Table 1).

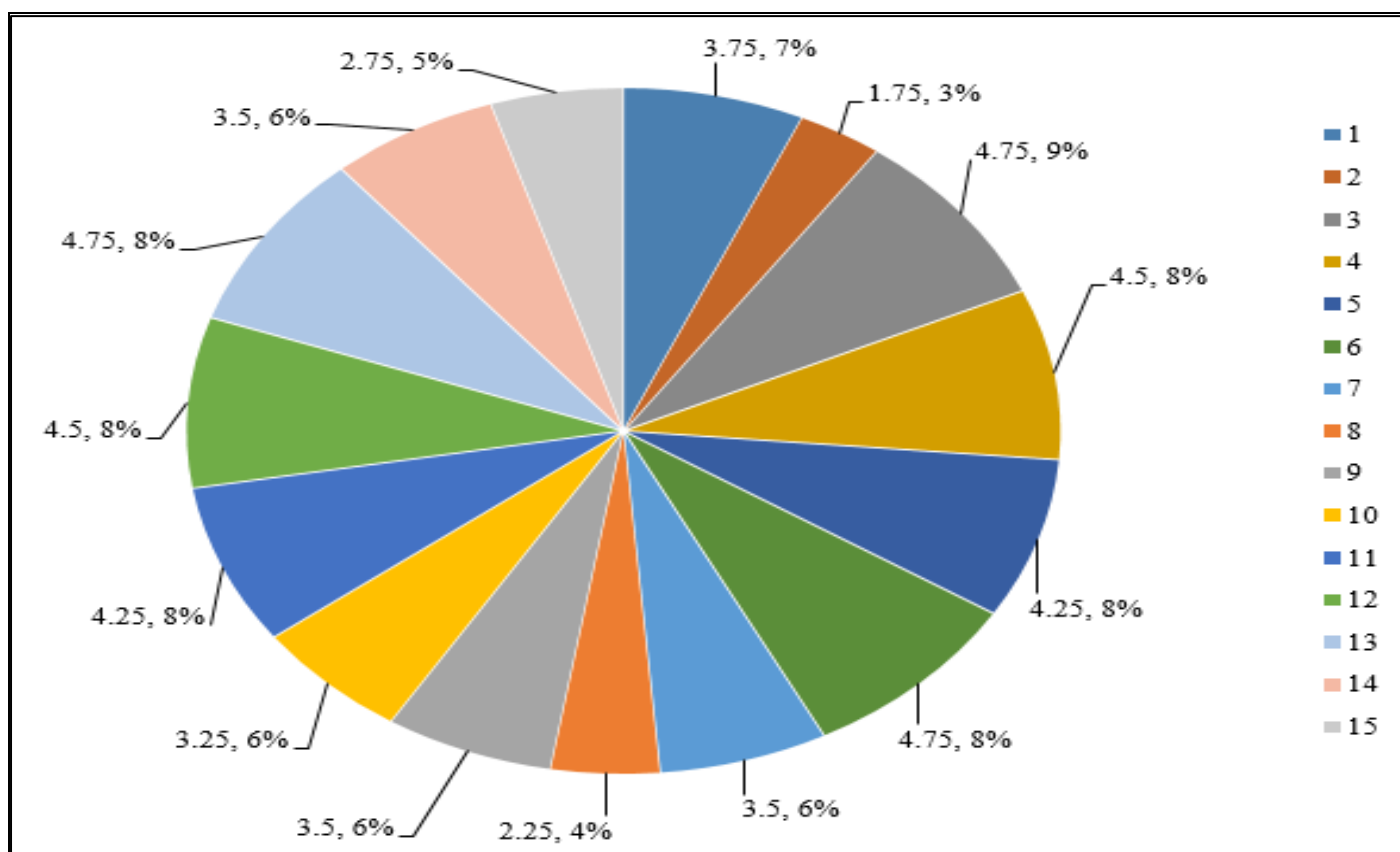


Fig 1: Mean Score Rating for Each Item of the Instrument

Table 1: The Pearson r Coefficients for the Items

Correlations					
		RATER1	RATER2	RATER3	RATER4
RATER1	Pearson Correlation	1	-.073	.227	.394
	Sig. (2-tailed)		.796	.415	.147
	N	15	15	15	15
RATER2	Pearson Correlation	-.073	1	.180	.223
	Sig. (2-tailed)	.796		.521	.424
	N	15	15	15	15
RATER3	Pearson Correlation	.227	.180	1	.317
	Sig. (2-tailed)	.415	.521		.250
	N	15	15	15	15
RATER4	Pearson Correlation	.394	.223	.317	1
	Sig. (2-tailed)	.147	.424	.250	
	N	15	15	15	15

The Pearson r coefficient from these data indicate an overall positive relationship between and among the rater's ratings except rater 1 and 2 (r-073).

Table 2: The Spearman Rho Coefficients for the Items

Correlations						
			RATER1	RATER2	RATER3	RATER4
Spearman's rho	RATER1	Correlation Coefficient	1.000	.003	.204	.314
		Sig. (2-tailed)	.	.993	.466	.254
		N	15	15	15	15
	RATER2	Correlation Coefficient	.003	1.000	.180	.113
		Sig. (2-tailed)	.993	.	.520	.689
		N	15	15	15	15
	RATER3	Correlation Coefficient	.204	.180	1.000	.230
		Sig. (2-tailed)	.466	.520	.	.409
		N	15	15	15	15
	RATER4	Correlation Coefficient	.314	.113	.230	1.000
		Sig. (2-tailed)	.254	.689	.409	.
		N	15	15	15	15

When the ratings are considered as ranks the speaman Rho relationship check indicated similar results: all ratings by 4 raters correlated positively but no negative relationship was

detected. The values were lower than the ones obtained in by the speaman.

➤ *The Interrater Reliability Cronbach Alpha Coefficients for the Items*

Table 3: Inter-Rater Reliability Cronbach's Alpha Coefficients for all Raters

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
RATER1	11.20	6.457	.242	.460
RATER2	11.60	4.829	.180	.553
RATER3	11.73	5.781	.352	.381
RATER4	12.67	3.810	.448	.221

➤ *The Intraclass Reliability Alpha Coefficients for the Items*

Correlation coefficients for the inter-rater reliability of the item ratings show..... as indicated in the table below (Table 2)

Given the weakness of the 2 reliability checks ( Speaman Rho and Pearson), no significance of relationship indicated in either procedure, the Cronback alpha reliability

was used. A cronback coefficient 0.70 was obtained dressing an acceptable level of reliability.

The intraclass correlation of 0.33 is low when a simple measure is considered.

The coefficient of . 71 is obtained when average measure is considered. Thus, one should rely on the combined ratings more than a single rater's rating of this instrument.

Table 4: Intraclass Correlation Coefficient

	Intraclass Correlation <sup>a</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.192 <sup>b</sup>	-.029	.512	1.950	14	42	.048
Average Measures	.487 <sup>c</sup>	-.126	.808	1.950	14	42	.048

#### IV. DISCUSSION AND CONCLUSION

Globally our hypothesis was rejected since the majority of the population rated the instrument invalid. This article focused on the assessment of the instrument for its validity to collect data on the topic “**STUDENTS LEARNING STYLES IN AN EFL CONTEXT: A CASE STUDY**”. The study shows limitation in that the quality check using quantitative measures of reliability indicate a low level of reliability for each rater and a low level of agreement among raters. The negative correlation between two of the raters raises some serious concerns leading to 2 possible explanations: either the items are ambiguous or the raters have not been able to rate the items using similar criteria for the rating or even their misunderstanding of the rating. Overall, from what our raters have said, we can only conclude that the instrument was valid and be administered to respondents for fine data collection on this issue.<sup>1</sup> This instrument evaluation form needs to be improved and submitted to another set of raters or else train the current raters in the approach to rating when using this instrument evaluation form to minimize their level of disagreement.

#### REFERENCES

- [1]. Barrett, Paul (2001). Assessing the Reliability of Rating Data
- [2]. Bresciani, Marilee J., Oakleaf, Megan, Kolkhorst, Fred, Nebeker, Camille, Barlow, Jessica, Duncan, Kristin, and Hickmott, Jessica (2009). Examining Design and Inter-Rater Reliability of a Rubric Measuring Research Quality across Multiple Disciplines. *Practical Assessment, Research & Evaluation*, 14(12). Available online: <http://pareonline.net/getvn.asp?v=14&n=12>.
- [3]. Brewer, B.W. (2002). Avondoglio JB, Cornelius A. E et al. Construct validity and interrater agreement of the Sport Injury Rehabilitation Adherence Scale. *Journal Sport and Rehabilitation* 11:170-178.
- [4]. Bryman, A. (2006): Integrating quantitative and qualitative research: how is it done? *Qualitative Research* doi:10.1177/1468794106058877
- [5]. Kimberlin Carole I. and almut Winterstein G. (2008). Validity and reliability of measurement instruments used in research. *Am J Health-Syst Pharm—Vol 65*
- [6]. Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20, 37 – 46.
- [7]. Gwet, L. (2008). Computing Inter-Rater Reliability and its Variance in the Presence of High Agreement. *British Journal of Mathematical and Statistical Psychology*, 61, 29 – 48.
- [8]. Gwet, L. (2010). *The Handbook of Inter-Rater Reliability*. Gaithersburg: Advanced Analytics.
- [9]. Hallgren, Kevin A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial Tutorials in Quantitative Methods for Psychology 2012, Vol. 8(1), p. 23-34.
- [10]. Havercamp, Susan Evaluating the Reliability and Validity of the Questionnaire for Situational Information: Test-Retest Reliability Final Report
- [11]. Inter-rater Agreement for Nominal/Categorical Ratings. <http://>\_\_\_\_\_
- [12]. Inter-rater Agreement for Ranked Categories of Ratings. <http://>\_\_\_\_\_
- [13]. Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. (Also available online at <http://aera.net/publications/?id=338>)
- [14]. Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- [15]. Kathy J. Rye, Commissioner Incorporating Inter-Rater Reliability into Your Curriculum. [www.coarc.com](http://www.coarc.com)
- [16]. Kilem Li Gwet (2008) Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* (2008), 61, 29–48
- [17]. Matthew Graham, Anthony Milanowski, Jackson Miller, & Westat (February 2012) Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. Center for Education Compensation Reform (CECRF)
- [18]. McCray, G. (2013) Assessing inter-rater agreement for nominal judgement variables. Paper presented at the Language Testing Forum. Nottingham, November 15-17.
- [19]. Rust, Roland (2001) Interrater Reliability. *Journal Of Consumer Psychology*, 10(1&2), 71–73
- [20]. STURGILL, LYNNE P., Lynn Snyder Mackler, Tara Jo Manal, & Michael J. Axe, (2009) Interrater Reliability of a Clinical Scale to Assess Knee Joint Effusion. *Journal of Orthopaedic & Sports Physical*

<sup>11</sup> Acknowledgement we are grateful to Professor Frank Buhendwa Munganga for his contribution and his availability without which this work would not come to end;

Mr. Banyongi Manyole Odida for his contribution in data analysis.

- Therapy , Volume 39 number 12 december 2009 pp. 845-849. Doi: 10. 2519/jospt.
- [21]. Walsh P, Thornton J, Asato J, Walker N, McCoy G, Baal J, Baal J, Mendoza N, Banimahd F. (2014) Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. PeerJ 2:e651 <https://dx.doi.org/10.7717/peerj.651>
- [22]. Zegers, Frits E. Coefficients for Interrater Agreement. Applied Psychological Measurement Vol. 15, No. 4, December 1991, Pp. 321-333
- [23]. Walsh P, Thornton J, Asato J, Walker N, McCoy G, Baal J, Baal J, Mendoza N, Banimahd F. (2014) Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. PeerJ 2:e651 <https://dx.doi.org/10.7717/peerj.651>.

**SURVEY QUESTIONNAIRE**

Dear Sir/Madam, this is a study on “**THE USE VIDEO EPISODES IN AN EFL CONTEXT: CASE OF FIRST YEAR AIA**”. For which we would like you to contribute favorably by answering the following 13 items for the sake of their validity.

- Note: Excellent =5, Very good = 4, Good =3, Poor = 2 and very poor = 1

Please, rate the items in this box.					
Excel	V.good	Good	poor	V.poor	
5	4	3	2	1	
<b>I. Identification of the participants:</b> Check the box that fits your identity 1. Sex: Male <input type="checkbox"/> Female <input type="checkbox"/> 2. Nationality: Congolese <input type="checkbox"/> Other <input type="checkbox"/> 3. Marital status: Married <input type="checkbox"/> Single <input type="checkbox"/>					
<b>II. Questions of the investigation</b> <b>1. Do your professors use video episodes in their teaching?</b> Yes, they do <input type="checkbox"/> No, they don't <input type="checkbox"/>					
<b>2. How many times a week do they use video episodes?</b> Once <input type="checkbox"/> Twice <input type="checkbox"/> Three times <input type="checkbox"/> Four times <input type="checkbox"/> Five times <input type="checkbox"/> Six times <input type="checkbox"/>					
<b>3. When do your teachers use the video episode?</b> At the beginning <input type="checkbox"/> during the lesson <input type="checkbox"/> <input type="checkbox"/> at the end of the lesson <input type="checkbox"/>					
<b>4. How do you feel when a teacher uses a video episode?</b> I feel happy <input type="checkbox"/> I feel sad <input type="checkbox"/> I feel frustrated <input type="checkbox"/> I don't come to the class <input type="checkbox"/>					
<b>5. Do these feelings change over time after you have undergone some English learning video episodes?</b> Yes, they do <input type="checkbox"/> No, they don't <input type="checkbox"/>					
<b>6. Do you use video episode after the class to learn English?</b> Yes, I do <input type="checkbox"/> No, I don't <input type="checkbox"/>					
<b>7. If yes, how many times a week do you use video episodes after class?</b> Once <input type="checkbox"/> Twice <input type="checkbox"/> Three times <input type="checkbox"/> Four times <input type="checkbox"/> Five times <input type="checkbox"/> Six times <input type="checkbox"/>					
<b>8. Is it good to learn English through a video episode?</b> Yes, it is <input type="checkbox"/> No, it is not <input type="checkbox"/>					
1					
2					
3					
1					
2					
3					
4					
5					
6					
	Excl	V.good	Good	Poor	V
5					
6					

<p><b>9. What are some advantages of learning with video episodes?</b></p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>	7					
	8					
	9					
		Excl	V.good	Good	Poor	V.poor
	10					
	11					
<b>12.a. Any item to improve? ( please arrange it here)</b>	12. a					
<b>12.b. Do you have any comment? (please do it here)</b>	12.b.					