



Assessing the Reliability and Validity of Research Instruments for Measuring Youth Purchase Intention in Xi'an, China: A Pilot Testing

Mo Jiao Jiao^{a*}, Chin Nyuk Sang^b

^{a,b}*Malaysia University of Science and Technology (MUST), Block B, Encorp Strand Garden Office, No. 12,
Jalan PJU 5/5, Kota Damansara, 47810 Petaling Jaya, Selangor, Malaysia*

^a*Email: mo.jiaojiao@phd.must.edu.my*

^b*Email: nschin@must.edu.my*

Abstract

The purpose of this pilot study is to evaluate the reliability and validity of research instruments developed to measure the purchase intention of youth in Xi'an, China. Given the importance of understanding youth purchase intentions and behavior for businesses and marketers, it is critical to have validated research tools, which is not currently the case. The purpose of this study is to develop and evaluate the reliability and validity of such research tools using a sample of young consumers in Xi'an. Overall, the combination of Cronbach's alpha, Pearson's correlation coefficient, construct validity, and Fleiss' Cohen's kappa coefficient are used to thoroughly test the reliability and validity of the research instruments in this study. The research design for this study is quantitative and uses a structured questionnaire administered to a sample of 50 young consumers aged 18 to 35 in Xi'an to investigate the factors that influence their purchase intentions in live streaming e-commerce. The results show that the research instruments are reliable and validated for the main data collection. The results of this study will contribute to the existing knowledge of young people's purchase intentions in China and provide valuable insights for companies and marketers.

Keywords: Pilot study; E-commerce; Chinese youth; Research instruments; Reliability; Validity.

Received: 5/10/2023

Accepted: 6/20/2023

Published: 6/30/2023

* Corresponding author.

1. Introduction

The year 2020 witnessed an unforeseen outbreak of COVID -19 that had a significant impact on various industries. The widespread adoption and rapid development of the Internet and 5G technology have led to increased use of mobile devices by people. Consequently, mobile short videos have become a very popular form of content. As of December 2020, there were 927 million online video users in China, accounting for 93.7% of Internet users. Among these users, 873 million used mobile short video platforms such as TikTok and Kuaishou, accounting for 88.3% of China's internet population. According to the 47th China Internet Growth Statistics Report (2021) [1], there were 388 million users who engaged in live streaming e-commerce in China, accounting for 39.2% of Internet users. Of these users, 66.2% made purchases while watching live streams. In 2019, the live streaming e-commerce industry in China recorded a total volume of 433.8 billion yuan, a year-on-year increase of 226%.

With the increasing trend of e-commerce and online shopping, live streaming e-commerce has become a popular method of online shopping. Live-streaming e-commerce involves promoting and selling products through live-streaming videos of influencers or sellers. Live streaming e-commerce provides an interactive and engaging shopping experience that can influence consumer purchase intent. Despite the increasing adoption of live streaming e-commerce, research on live streaming shopping is still at an early stage, especially in China. While some scholars have recognized the importance of the primary live streamer, also known as the "anchor," as a key aspect of live streaming e-commerce and have studied how the characteristics and attributes of the anchor influence consumers' internal state and online purchase behavior, research in this area is still limited and there is lack of research instruments to be applied in this study. Therefore, the study aims to test the validity and reliability of the research procedures and instruments used in the main study. The study also aims to identify any weaknesses in the research design and improve the effectiveness and quality of the main study.

2. Methodology

Validation of the research instrument is a critical step to ensure the reliability and validity of the data collected. Before conducting the pilot study, the questionnaire measuring the factors influencing Chinese youths' purchase intention in live streaming e-commerce will be sent to three experts for content analysis and review. These experts will thoroughly evaluate the questionnaire for relevance, clarity, and completeness in measuring the intended constructs or variables of interest. All feedback received from the experts will be carefully considered and necessary changes will be made to improve the questionnaire.

After content analysis and review by the experts, the questionnaire will be subjected to a pilot test with a sample of 30 Chinese adolescents from Xi'an. The main goal of the pilot test is to evaluate the reliability and validity of the questionnaire on a smaller sample before conducting the main study with a larger sample.

During the pilot test, the 50 Chinese youths will be provided with the questionnaire to complete, which will measure their intention to purchase live streaming e-commerce. The data collected during the pilot testing will be carefully analyzed to evaluate the reliability and validity of the questionnaire. Reliability will be assessed by

examining the internal consistency of questionnaire items using techniques such as Cronbach's alpha, which measures the extent to which items in a scale or questionnaire are correlated with each other. Construct validity is assessed by examining the extent to which the questionnaire measures the intended constructs or variables of interest using techniques such as factor analysis, which examines the underlying factor structure of questionnaire items.

Based on the results of the pilot testing, any necessary changes or modifications are made to the questionnaire to improve its reliability and validity. This may include revising or rewording certain items, deleting irrelevant items, or adding new items based on feedback from pilot test participants. The finalized and validated questionnaire will then be used in the main study with a larger sample to collect data for analyzing the factors influencing Chinese youths' purchase intention in live streaming e-commerce in Shanxi province.

Table 1: Pilot Test Procedures.

Phase	Description
Step 1	Content analysis and review of the questionnaire by three experts in the field for relevance, clarity, and completeness
Step 2	Incorporation of feedback received from experts to improve the questionnaire
Step 3	Pilot test of the questionnaire with a sample of 50 Chinese adolescents from Xi'an
Step 4	Analysis of data collected during pilot testing to evaluate reliability and validity using techniques such as Cronbach's alpha for internal consistency and factor analysis for construct validity
Step 5	Modification of the questionnaire based on results of pilot testing to improve reliability and validity, including revising or rewording items, deleting irrelevant items, or adding new items
Step 6	Finalization and validation of the questionnaire for use in the main study with a larger sample in Shanxi province

2.1. Research Instruments

The questionnaire was adapted from two validated resources (See Table 2).

Table 2: Research Questionnaires Adapted from Two Validated Resources.

Authors	Research Title	Variables
Yan, Tianmeng (2022) [2]	The Impact of Live Streaming E-commerce Features on Chinese Consumers' Attitude and Purchase Intention" (2022).	<ul style="list-style-type: none"> • Visibility • Interactivity • Influencer Endorsement • Perceived behavior control • Subjective Norms • Attitude • Purchase intention
Zhou, Rong & Tong, Lei. (2022) [3]	A Study on the Influencing Factors of Consumers' Purchase Intention During Livestreaming E-Commerce: The Mediating Effect of Emotion.	Emotional appeal

The questionnaire was designed in the following steps. The first step is to determine the research goals and objectives. This survey was designed according to the type of respondents. The researcher reviewed the

objectives, purpose, research questions, theoretical and conceptual framework of the study to design the current research questionnaire [4].

In the second step, after designing the research questionnaire, the target subject was identified. The relevant information for the questionnaire can only be collected if the appropriate respondents are selected. The statements in this questionnaire were designed according to the nature and mental level of the respondents so that they can easily understand and answer them. The characteristics of the group were almost the same. It is impossible to design a single questionnaire when the selected sample is a heterogeneous group [5].

In the third step, the statements in the questionnaire were designed. These statements were worded according to the type of respondents. The statements were reworded and validated using expert opinions to make them meaningful so that appropriate responses were obtained from students [6]. In the fourth step, the researcher decided to develop closed-ended questions. These closed-ended questions were designed on a Likert scale (strongly agree, agree, neutral, disagree, and strongly disagree). The closed-ended questions take less time and respondents can answer them easily. The closed-ended questions were developed because they are considered appropriate for quantitative research [7].

In the final step, this questionnaire was distributed to 50 adolescents in Xi'an. The data collected from these participants were entered into the Statistical Package for Social Sciences (SPSS). Cronbach's Alpha reliability coefficient was used to check the reliability of the questionnaire.

The questionnaires were divided into seven sections according to the research objectives, as shown in Table 3.2. The proportions of the samples and their sizes in the study region are shown in the table below. The Mandarin version of the research instruments was translated using the Brislin method, which involves translating a previously translated text back into the source language, as recommended by [8] & [9] for evaluating translation work. In this study, the researcher evaluated the reliability of the Mandarin version of the research instruments, and the results showed a Cronbach's alpha value of .73, which exceeds the acceptable threshold of .70, indicating good reliability.

The research instruments used in this study include a questionnaire designed to measure the purchase intention of young people in Xi'an. The questionnaire is developed based on relevant literature and contains items on various factors that influence adolescents' purchase intention, such as visibility, influencer endorsement, information display, interactivity, attitude, social norm, perceived behavior control.

3. Pilot Study Findings & Results

Descriptive statistics are used to summarize participant demographic characteristics. Reliability analysis using Cronbach's alpha will be conducted to assess the internal consistency of the research instruments. Exploratory factor analysis will be conducted to examine the factor structure of the questionnaire and assess construct validity.

3.1. Descriptive Analysis

Table 3 shows that of the 50 respondents, 27 identified themselves as male and 23 as female, representing 54% and 46% of the sample, respectively. This indicates a slightly higher proportion of males in the study.

Table 3: Descriptive based on Gender.

Gender	Frequency (n)	Percentage (%)
Male	27	54%
Female	23	46%
Total	50	100

Source: Researcher ‘s data collection, April 2023

The age groups mentioned in the Table 4 are: 18-21 years old, 22-25 years old, 26-29 years old, 30-33 years old, and 34-35 years old. 8 subjects (16.0% of the total sample) are 18-21 years old, 13 subjects (26.0% of the total sample) are 22-25 years old, 9 subjects (18.0% of the total sample) are 26-29 years old, 10 subjects (20.0% of the total sample) are 30-33 years old, and 10 subjects (20.0% of the total sample) are 34-35 years old.

Table 4: Descriptive based on Age.

Gender	Frequency (n)	Percentage (%)
18-21 years old	8	16.0%
22 - 25 years old	13	26.0%
26 - 29 years old	9	18.0%
30 - 33 years old	10	20.0%
34 – 35 years old	10	20.0%
Total	50	100%

3.2. Reliability Testing

Prior to conducting data analysis, it is essential to evaluate the validity and reliability of the measurement tool utilized, which in this case is the questionnaire. The evaluation of the questionnaire's validity and reliability was conducted using the IBM SPSS Statistics version 27 software. Specifically, Cronbach's Alpha was employed to measure the reliability of the questionnaire, yielding an overall result of 0.915. This value surpasses the minimum requirement of 0.70, as recommended by [10], indicating a high level of reliability for the questionnaire. These findings are sufficient for evaluating the convergent validity of the questionnaire. Table 5 displays the results of the reliability testing for all scales.

The variables "V," "EA", "IE", "A," "SN", "PBC," "PI", "I," and "ID" have Cronbach's alpha values higher than the generally accepted minimum value of 0.70, indicating high reliability.

These variables have Cronbach's alpha values of 0.905, 0.902, 0.886, 0.918, 0.894, 0.914, and 0.904, 0.886, and 0.928, respectively.

Table 5: Reliability for All Constructs.

Variables	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
V	136.18	920.436	.699	.905
I	135.80	851.184	.959	.886
EA	136.18	907.987	.744	.902
IE	135.80	851.184	.959	.886
ID	135.66	1036.678	.341	.928
A	135.82	1008.681	.486	.918
SN	135.82	903.375	.864	.894
PBC	135.72	971.389	.569	.914
PI	136.86	904.368	.716	.904

3.3. Validity Testing

Table 5 contains correlation coefficients (r-values) for various variables indicating the strength of their association or relationship. These coefficients can provide information about the validity of the variables in the context of the study.

In general, higher correlation coefficients indicate stronger relationships between variables, while lower coefficients indicate weaker relationships.

For example, a correlation coefficient of 1 represents a perfect positive relationship, -1 represents a perfect negative relationship, and 0 represents no relationship.

Table 5 shows that some variables have moderate to high positive correlations with each other, as indicated by correlation coefficients close to 1. For example, the variables "I" and "IE" have a correlation coefficient of 1, indicating a perfect positive relationship. Similarly, "I" and "SN" have a high positive correlation coefficient of 0.901, indicating a strong relationship between them.

On the other hand, some variables have weaker relationships with each other, indicated by lower correlation coefficients. For example, "ID" and "PBC" have a correlation coefficient of -0.028, indicating a weak negative relationship between them.

Correlation coefficients do not imply a causal relationship, and the validity of variables cannot be determined based on correlations alone. Validity refers to the accuracy with which a measurement instrument or variable measures what it is intended to measure.

While correlation coefficients can provide evidence of convergent validity, which assesses the positive relationship between different measures of the same construct, validity should be assessed using a comprehensive approach that considers multiple sources of evidence, such as content validity, criterion validity, and construct validity.

To assess the validity of the research instruments, we sent the questionnaires to three experts for review. Section 3.4 discuss about content validity.

Table 5: Validity Testing.

Correlations							
		V	I	EA	IE	ID	total_A
V	Pearson Correlation	1	.678**	.788**	.678**	.117	.264
	Sig. (2-tailed)		.000	.000	.000	.417	.064
	N	50	50	50	50	50	50
I	Pearson Correlation	.678**	1	.727**	1.000**	.463**	.540**
	Sig. (2-tailed)	.000		.000	.000	.001	.000
	N	50	50	50	50	50	50
EA	Pearson Correlation	.788**	.727**	1	.727**	.170	.308*
	Sig. (2-tailed)	.000	.000		.000	.238	.029
	N	50	50	50	50	50	50
IE	Pearson Correlation	.678**	1.000**	.727**	1	.463**	.540**
	Sig. (2-tailed)	.000	.000	.000		.001	.000
	N	50	50	50	50	50	50
ID	Pearson Correlation	.117	.463**	.170	.463**	1	.238
	Sig. (2-tailed)	.417	.001	.238	.001		.096
	N	50	50	50	50	50	50
A	Pearson Correlation	.264	.540**	.308*	.540**	.238	1
	Sig. (2-tailed)	.064	.000	.029	.000	.096	
	N	50	50	50	50	50	50
SN	Pearson Correlation	.542**	.901**	.597**	.901**	.531**	.644**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	50	50	50	50	50	50
PBC	Pearson Correlation	.670**	.588**	.632**	.588**	-.028	.253
	Sig. (2-tailed)	.000	.000	.000	.000	.845	.077
	N	50	50	50	50	50	50
PI	Pearson Correlation	.540**	.793**	.579**	.793**	.272	.324*
	Sig. (2-tailed)	.000	.000	.000	.000	.056	.022
	N	50	50	50	50	50	50
**. Correlation is significant at the 0.01 level (2-tailed).							
*. Correlation is significant at the 0.05 level (2-tailed).							

3.4. Content Validity

Cohen's Kappa is a statistical measure for assessing the agreement between two raters, taking into account the agreement due to chance. However, because this study involves more than two raters, Fleiss' kappa was used instead (Refer Table 6). Fleiss' Kappa is a statistical measure that falls within the range of -1 to +1. A negative value for kappa (κ) indicates that the agreement between two or more raters was lower than expected by chance. A value of -1 suggests no observed agreement, meaning the raters did not agree on anything. A value of 0 (zero) indicates that the agreement was no better than chance, although negative values are rare in practice [11]. On the other hand, kappa values greater than 0 (zero) indicate increasing levels of agreement that surpass chance, with a maximum value of +1 representing perfect agreement, where the raters agreed on everything. Table 6 shows the Fleiss' kappa. In this case, the kappa value is reported as 0.630, indicating the overall agreement among the three raters. Thus, all of the constructs in this research questionnaires are validated and reliable.

Table 6: Fleiss' Kappa Test.

Overall Agreement^a						
	Kappa	Asymptotic			Asymptotic 95% Confidence Interval	
		Standard Error	z	Sig.	Lower Bound	Upper Bound
Overall Agreement	.630	.086	7.316	.000	.461	.798
a. Sample data contains 45 effective subjects and 3 raters.						

4. Discussion

Feedback from the three experts on the research tools is generally positive and highlights the strengths of the tools. Expert 1 noted that the items are clearly and concisely worded, making them easy for the target audience to understand. The instruments were also developed based on established theoretical frameworks and previous research literature, indicating a solid foundation for the study. In addition, Expert 1 mentioned that the research instruments have good construct validity because they comprehensively measure the intended constructs or variables of interest.

Expert 2 emphasized that the response options and scales used in the research instruments are appropriate and well-defined, allowing for meaningful data collection and analysis. The instruments also appear to be culturally sensitive, taking into account the cultural context of the study population and ensuring relevance and appropriateness to the intended participants.

In addition, Expert 3 highlighted the good reliability of the research instruments as evidenced by the internal consistency and test-retest reliability analysis conducted. However, Expert 3 also suggested that providing additional information about the validation process, such as pilot testing, expert review, and changes made in response to feedback, would further strengthen the validity of the research instruments.

To improve the research tools based on the feedback, it would be beneficial to consider the following suggestions:

1. Provide additional information about the validation process by including details about the pilot testing, expert review, and changes made based on feedback can increase the transparency and credibility of the research instruments. This information would demonstrate that the instruments have undergone a rigorous validation process, increasing their reliability and validity.
2. Consider incorporating expert suggestions: Expert feedback can be carefully reviewed and considered for possible changes to the research instruments. For example, if Expert 2 suggests that the response options and scales are appropriate, ensure that they are consistently applied to all items and are consistent with the research objectives.
3. Cultural Sensitivity Review: although Expert 2 mentioned that the research instruments appear to be culturally sensitive, it would be worthwhile to conduct a thorough review to ensure that the items are relevant and appropriate for intended participants from different cultural backgrounds. This could include consulting with cultural experts or conducting pilot testing with participants from different cultural groups.
4. Check whether item wording is ambiguous or complex: Review the items in the research instruments to ensure

that they are clearly and concisely worded, as noted by Expert 1. Avoid using jargon or complex language that may confuse the target audience, and consider conducting pilot tests to gather feedback on item clarity.

5. Assess and continuously ensure reliability: As mentioned by Expert 3, reliability of research instruments is critical. Periodically throughout the study, assess the internal consistency and test-retest reliability of the instruments to ensure that they are providing consistent and reliable data.

Incorporating these suggestions and considering expert feedback can further improve the research instruments, leading to increased validity and reliability and ultimately increasing the quality and credibility of the research findings.

5. Conclusion

As a conclusion, this pilot study is expected to provide valuable insights into the factors influencing Chinese youths' purchase intention in live streaming e-commerce in Shanxi Province.

Reliability and validity testing of the questionnaire used in the study will ensure the robustness of the research findings. By assessing the internal consistency of the items using Cronbach's alpha coefficient for reliability and conducting content validity and construct validity tests, the study will ensure that the questionnaire is a reliable and valid instrument to measure the intended constructs.

However, this pilot study has some limitations. The small sample size of 50 respondents may limit the generalizability of the results to a larger population. In addition, the cross-sectional design, in which data are collected at a specific point in time, may not capture changes in consumer behavior over time. These limitations should be considered when interpreting the results of the pilot study.

Therefore, in order to gain a more comprehensive understanding of the factors influencing Chinese youths' purchase intention of live streaming e-commerce in Shanxi Province, further research with a larger sample and a longitudinal design in which data are collected over a longer period of time is recommended. This can help validate and extend the findings of the pilot study and provide more informed insights for businesses and policymakers to make informed decisions.

References

- [1] China Internet Network Information Center (CNNIC), "The 47th China Statistical Report on Internet Development," 2021. [Online]. Available: http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm. [Accessed: Insert Date Accessed].
- [2] T. Yan, "The Impact of Live Streaming E-commerce Features on Chinese Consumers' Attitude and Purchase Intention," Theses - ALL, 643, 2022. [Online]. Available: <https://surface.syr.edu/thesis/643>. [Accessed: Insert Date Accessed].
- [3] R. Zhou and L. Tong, "A Study on the Influencing Factors of Consumers' Purchase Intention During Livestreaming e-Commerce: The Mediating Effect of Emotion," *Frontiers in Psychology*, vol. 13,

2022, doi: 10.3389/fpsyg.2022.903023.

- [4] S. Yaddanapudi and L. N. Yaddanapudi, "How to design a questionnaire," *Indian Journal of Anaesthesia*, vol. 63, pp. 335-337, 2019, doi: 10.4103/ija.IJA-334-19.
- [5] P. C. Beatty, D. Collins, L. Kaye, J. L. Padilla, G. B. Willis, and A. Wilmot, Eds., *Advances in Questionnaire Design, Development, Evaluation and Testing*, John Wiley & Sons, 2019, ISBN: 978-1-119-26362-3.
- [6] J. K. Boparai, S. Singh, and P. Kathuria, "How to Design and Validate A Questionnaire: A Guide," *Current Clinical Pharmacology*, vol. 13, no. 4, pp. 210-215, 2018, doi: 10.2174/1574884713666180807151328.
- [7] S. O'Connor, "Designing and Using Surveys in Nursing Research: A Contemporary Discussion," *Clinical Nursing Research*, vol. 31, no. 4, pp. 567-570, 2022, doi: 10.1177/10547738211064739.
- [8] Jones, Patricia & Lee, Jerry & Phillips, Linda & Zhang, Xinwei & Jaceldo, Karen. (2001). An Adaptation of Brislin's Translation Model for Cross-cultural Research. *Nursing Research - NURS RES.* 50. 300-304. 10.1097/00006199-200109000-00008.
- [9] Brislin, R. W., "The wording and translation of research instruments," in W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research*, pp. 137-164, Beverly Hills, CA: Sage, 1986. [IEEE citation]
- [10] Pallant, J., "SPSS kullanma kılavuzu SPSS ile adım adım veri analizi (Çev: Balcı, S. & Ahi, B.)," Ankara: Anı Yayıncılık, 2016.
- [11] Agresti, A., "Analysis of ordinal categorical data," Vol. 656, John Wiley & Sons, 2010.