

CHAPTER 3 OBJECT AND RESEARCH METHODOLOGY

3.1 Object on The Research

This study takes short-video content marketing as the research object, and the respondents involved in short-video content marketing as the survey object. This study investigates the relationship between interactive content, product content, and entertainment content, and uses a quantitative approach to investigate the mechanism of short video content marketing on purchase intention from a theoretical perspective. This study consists of independent variables, namely content marketing (X) with sub-variables product content (X1), entertainment content (X2) and interactive content (X3), and the dependent variable is purchase intention (Y).

Respondents in this study are Three Squirrels (@songshu0619) TikTok followers in China. This study used a cross sectional study because data collection was only done once at a certain time (Siyoto, 2015). The research data collection period was carried out for less than one year, from April to July 2022.

3.2 Research Method

3.2.1 Types of research and Method

This study mainly uses the literature analysis method, questionnaire survey method and quantitative analysis method to study the theoretical framework.

1. Literature analysis. Based on reading, sorting out, and summarizing relevant literature and materials, the relationship between each construct is clarified, the theoretical model is constructed, and this study's hypothesis is further proposed (Suh & Prophet, 2018).
2. Questionnaire survey. By referring to existing studies, the mature scale developed by predecessors with high reliability and validity was selected and integrated into a questionnaire on this basis to complete the measurement of constructs and provide support for the verification of theoretical hypotheses later (Peña et al., 2002).
3. Quantitative analysis. SPSSAU were used to process and analyze the collected data, and descriptive statistics, exploratory factor analysis, reliability and validity detection, correlation analysis, hypothesis testing and other operations were

completed accordingly. Examine the hypothetical relationship between interactive Content, Entertainment Content, product Content, and purchase intention to reach accurate conclusions (Tárnok, 2014).

3.2.2 Operational Variables

Operational variables are the process of changing or decomposing concepts or constructs into measurable variables suitable for testing (Cooper & Schindler, 2014). This study consists of the independent variable, namely interactive content (X1), entertainment content (X2) and product content (X3), the dependent variable, namely purchase intention (Y).

Table 3.1
Operational Variables

Variable	Sub Variables/D imensions	Concept	Indicato r	Size	Scale	No. item
Content Marketing (X)	Content marketing is to deliver valuable and entertaining product or brand information through various channels and various forms of media content, to trigger customer participation, and establish and improve brand marketing strategy in the process of interaction (Chen & Zhou, 2013)					
	Interactive content (X1)	Interactive content is to realize consumers' emotional appeal, communication and information exchange and sharing needs (Xu Rujing 2018).	Sociality	the level of sociality of interactive content in short video	Interval	1
			distance	the level of distance between user and creator	Interval	2
			similar	the level which makes user meet the people who similar with themselves	Interval	3
			Communication	the level of communication of interactive content in short video	Interval	4
			companionship	the level of companionship of interactive content in short video	Interval	5
Entertainment Content (X2)	Entertainment content refers to the content that can attract users and	Amusement	the level which makes users amused by entertainment content in the short video	Interval	6	

Variable	Sub Variables/Dimensions	Concept	Indicator	Size	Scale	No. item
		generate emotional resonance (Lieb R, 2012).	excitement	the level which makes users excited by entertainment content in the short video	Interval	7
			imagination	the level of 33magination of entertainment content in the short video	Interval	8
			comfort	the level which makes user comfortable by entertainment content in the short video	Interval	9
			Addiction	the level which makes users addicted to entertainment content in the short video	Interval	10
	Product Content (X3)	Product content refers to valuable content with educational significance and high quality that can help consumers make purchase decisions (Sun Tianxu, 2016).	Brand	the level which makes users feel the brand by product content in the short video	Interval	11
			instructions	the level which makes users learn how to use the product by product content in the short video	Interval	12
			price	the level which makes users know about the price by product content in the short video	Interval	13
			value	the level which makes users get the value information by product content in the short video	Interval	14
			quality	the level which makes users feel the quality by product content in the short video	Interval	15

Variable	Sub Variables/Dimensions	Concept	Indicator	Size	Scale	No. item
Purchase Intention (Y)		purchase intention is a subjective intention to purchase products or services, and purchase behaviour can be effectively predicted through the expression of intention to a certain extent.	Interest preferential	Interest level prioritize the product in the video compared with other similar product	Interval	16
			attention	level of attention to the product in the product	Interval	17
			Interest to buy	Level of Interest to buy the product in the short video	Interval	18
			Perceived Value	Level of Perceived Value about product	Interval	19
			recomendation	Level of recommending the product in the short video to other people	Interval	20

Source: The results of data processing

3.2.3 Types and Sources of Data

The data used in this study consisted of primary data and secondary data. Primary data is data from the original or first source. In contrast, secondary data is a historical data structure of the variables that have been collected and gathered before (McDaniel & Roger, 2013). The following is an explanation of primary and secondary data:

1. Primary Data

According to McDaniel and Gates (2015), primary data is new data collected to help solve problems in research. Primary data collection techniques are actively carried out by distributing questionnaires to a number of respondents in accordance with the target which is considered to be able to represent the entire population of research data. Based on this understanding of population, Respondents of this study are Three Squirrels (@songshu0619) TikTok followers in China.

2. Secondary Data

According to McDaniel and Gates (2015), Secondary data is obtained or collected from existing sources, usually from the library or previous research

reports. In this study, secondary data came from literature studies through various journals, marketing magazine articles, and articles took on the internet served.

Table 3.2
Types and Data Sources

No.	Data	Data Types	Data Sources
1	The short video content type from january to march in 2022	Primary Data	Three Squirrels TikTok account
2	Respondents' responses regarding characteristics	Primary Data	Data analysis results of Three Squirrels (@songshu0619) TikTok followers regarding characteristics
3	Respondents' responses regarding content marketing	Primary Data	Data analysis results of Three Squirrels (@songshu0619) TikTok followers regarding content marketing
4	Respondents' responses regarding purchase intention	Primary Data	Data analysis results of Three Squirrels (@songshu0619) TikTok followers regarding purchase intention
5	Graphic represent the growth of netizen size and internet penetration rate 2018-2021	Secondary Data	China Internet Network Information Center, CNNIC
6	Short video industry penetration in china by the end of 2020	Secondary Data	Quest Mobile data
7	The marketing share of e-commerce platform 2017-2021	Secondary Data	www.stock.pingan.com
8	Gross merchandise value (gmv) of tiktok event point	Secondary Data	www.chanmama.com
9	The price range of top 5000 sales product in the key event 2020-2021	Secondary Data	www.chanmama.com
10	Top 5 sales category in tiktok china in 2021	Secondary Data	www.chanmama.com

Source: Based on various sources (Data processed by researchers, 2022)

3.2.4 Population, Sample, and Sampling Technique

3.2.4.1 Population

Determining the characteristics of the population which are elements in the object of research is an important step in collecting and analyzing data, in conducting research. The population is defined as a set of complete analysis units that are being

studied. The population relates to all groups of people, events, or objects that are the center of the attention of researchers to be studied (McDaniel & Gates, 2015). The population needs to be identified appropriately and accurately since the beginning of the research. If the population is not well identified, it will result in a mistaken conclusion of the research. Based on the understanding of the population, the population in this study are 5.16 million Three Squirrels (@songshu0619) TikTok followers in China (accessed from April to May 2022)

3.2.4.2 Sample

Samples are sub from a set of elements selected to be studied from the population (Blaikie, 2018). By taking samples, researchers can draw conclusions that will be generalized to the population. The population object is allowed to be taken from a portion of the specified amount, with the note that the part taken represents the others who are not examined (Bruce et al., 2020). A study cannot be a whole population studied. Therefore, researchers are allowed to take part of the population object determined by the record of the section that is taken represents that are not studied or representative (Nikam, 2018).

The author cannot study all populations in this study due to several factors, including limited costs, energy, and time available. Researchers are allowed to take part of the specified population object, with the note the part taken represents other population objects that are not examined. But the sample token for a study should represent the population. It's an important indicator for designing the testing sample design (Sekaran & Bougie, 2016).

According to Schumacke & Gorsuch (1996), there is a certain relationship between selecting sample size and the number of variables, and the proportion between them should be more significant than 1:5. In general, the ideal sample size should be 10-25 times the number of variables, and the overall sample size should not be less than 100, and the more the better. Rigdon (2005) believed that if the number of variables was more than 10 and the total number of samples was less than 200, the statistical analysis results obtained were not stable and convincing. In this paper, 20 variable factors are set up to study the relationship between content marketing and

consumers' purchase intention. Therefore, the sample of this survey should not be less than 200, considering that the research problem in this paper involves a wide range of regions and a large base of target consumers. At the same time, in the process of literature research, it was found that the sample number of the questionnaire survey conducted by most scholars on similar issues generally fell within the range of 200-270. Therefore, to make the accuracy of the questionnaire survey in this paper higher than the average level, the sample size of this questionnaire survey was collected 309.

3.2.4.3 Techniques Sampling

Sampling is the process of selecting the right number of elements from the population, so the sample of the research and an understanding of the characteristics allows us to generalize these properties or features in the population element (Sekaran & Bougie, 2016). There are two type of sampling technique:

Probability sampling is a sampling technique, also known as random sampling, in which each sampling unit in the overall population has the same probability of being selected for the sample. Random sampling has a sound basis in statistical theory, can be explained by probability theory, is an objective and scientific sampling method. In general, the basic form of probability sampling survey is divided into two categories: single-stage sampling and multi-stage sampling. Single-stage sampling refers to the sampling process only once, it has the following four kinds: simple random sampling, equidistant sampling, stratified sampling, whole-group sampling. Multi-stage sampling is the process of stratifying the total and then taking samples one by one. Multi-stage sampling is used when the overall population is particularly large. Regardless of the type of random sampling, the sample must always be viewed as an approximation of the total rather than the total itself (Sekaran & Bougie, 2016).

Non-probability sampling is a sampling method in which not all members of the population have an equal chance of participating in the study, unlike probability sampling. Each member of the population has a known chance of being selected. Non-probability sampling is most useful for exploratory studies like a pilot survey (deploying a survey to a smaller sample compared to pre-determined sample size).

Researchers use this method in studies where it is impossible to draw random probability sampling due to time or cost considerations.

Non-probability sampling is the sampling technique that is used in this study. Purposive sampling is used in this study, which does not provide the same opportunity for each element or member of the population to be selected as a sample. The researcher chooses the sample based on whom they think would be appropriate for the study. The sample in this study are Three Squirrels (@songshu0619) TikTok followers in China who are active in their comments and live streaming.

3.3 Data Collection Technique

According to Sekaran and Bougie, data collection methods are stages in critical research processes, because by getting the right data, the research process will take place until the researcher finds the answer to the formulation of a predetermined problem. In order to obtain data related to the subject under study, the authors will collect data by means of literature research and questionnaires:

1. Questionnaires

This study adopts the questionnaire survey method, using the Questionnaire Star, with the help of WeChat, TikTok and other social applications to send each people individually. Data collection always follows the policy of confidentiality, and Respondents in this study are Three Squirrels (@songshu0619) TikTok followers in China. Before the formal questionnaire survey, the research on interactive content, Entertainment Content, product content, and purchase intention should be collected and sorted. Refer to appropriate operational definitions and measurement scales, and then a more valid questionnaire was designed based on the maturity scale.

The questionnaire includes three parts: guideline, consumer purchase intention survey and personal characteristic information. To ensure the accuracy and effectiveness of the questionnaire data collection, keywords were indicated in the questionnaire to ensure the correct understanding of respondents, and basic information of respondents was collected to provide support for subsequent data analysis.

2. Literature study

The literature study is a collection of information related to the theories that have something to do with the problems and variables studied, consisting of literature studies on content marketing, purchase intention. The literature study was obtained from various sources, namely: a) Library of Indonesian Education University (UPI), B) Thesis, C) Journal of Economics and Business, D) Print Media (Magazines and Newspapers), E) Google Scholar Search engine, F) Science Direct Journal Portal, G) China National Knowledge Infrastructure (CNKI) and H) Emerald Insight journal portal and Elsevier Journal Portal.

3.4 Validity and Reliability Testing

From the scientific perspective of the study, the scale of the questionnaire should have adequate reliability and validity. Reliability examines the reliability of the questionnaire measurement, that is, the degree of internal consistency of the results obtained from the measurement of that scale. Validity refers to the accuracy of the scale measurement results, that is, the degree to which the results reflect the true situation. The process of analyzing structural equation models usually involves two stages: firstly, assessing the reliability and validity of the measurement model, and secondly, assessing the causal relationships within the structural model (Anderson & Gerbing, 1984). The two-stage approach ensures that the validated variables' relationships are based on the premise that all variables are measured credibly and are valid.

As of June 12, 2022, a total of 316 answers were collected. After excluding 12 of the invalid answers, 309 valid questionnaires were finally obtained, and the questionnaire was about 97%. The following will focus on the data analysis of the 309 questionnaire results.

3.4.1 Reliability Testing

The main reference values for reliability analysis were Correlated-Item Total Correlation (CITC) analysis and purified measurement terms. The exclusion of items with low correlation coefficients of single item total scores can improve the internal consistency reliability of the questionnaire. Generally, the item should be deleted when the Corrected item-total Correlation is less than 0.4 (Kerlinger, 1979). In addition, if

Alpha if Item Deleted increases compared to the original Cronbach's alpha coefficient, the questionnaire item should also be deleted. For the treatment of items with CITC values less than 0.3, there is an option to delete them to increase the alpha value. Reliability was measured using Cronbach's alpha, where a Cronbach's alpha greater than 0.7 indicates a satisfactory level of item reliability (Brannick & Zhang 2013).

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N - 1)\bar{c}}$$

Information:

α = instrument reliability

N = the number of items,

\bar{c} = the average inter-item covariance among the items and

\bar{v} = the average variance.

Table 3.3

Reliability Statistics (Cronbach Alpha)

Items	Corrected Item-Total Correlation(CITC)□	Cronbach Alpha if Item Deleted□	Cronbach α □
Q1	0.661	0.947	0.95
Q2	0.652	0.947	
Q3	0.639	0.948	
Q4	0.655	0.947	
Q5	0.628	0.948	
Q6	0.638	0.948	
Q7	0.62	0.948	
Q8	0.629	0.948	
Q9	0.665	0.947	
Q10	0.599	0.948	
Q11	0.675	0.947	
Q12	0.653	0.947	
Q13	0.674	0.947	
Q14	0.687	0.947	
Q15	0.684	0.947	
Q16	0.742	0.946	
Q17	0.812	0.945	
Q18	0.802	0.946	
Q19	0.796	0.945	

Q20	0.767	0.946	
-----	-------	-------	--

Source: Processed data (2022)

From the Table 3.3 and 3.4, it can be seen that the reliability coefficient value is 0.950, more significant than 0.9, indicating that the reliability quality of the research data is very high. For the " α coefficient of the deleted item," the reliability coefficient does not increase significantly after any item is deleted, indicating that the item should not be deleted.

For the "CITC value," the CITC values of the analysis items are all greater than 0.4, indicating a good correlation between the analysis items and a good reliability level. In summary, the reliability coefficient value of the study data is higher than 0.9, which indicates that the data reliability is of high quality and can be used for further analysis.

Table 3.4
Cronbach Alpha

N of Items	n	Cronbach α
20	309	0.95

Source: Processed data (2022)

3.4.2 Validity Testing

The reliability of the questionnaire is qualified, but its validity may not be qualified either. Reliability is a necessary but not sufficient condition for validity. Validity tests are the most valid. Validity refers to the degree to which a measurement instrument can correctly measure the property intended to be measured, that is, the effectiveness of the measurement.

Validity mainly includes content validity and structural validity. Content validity, also known as surface or logical validity, is a subjective and systematic assessment of the pros and cons of a measurement task in terms of the content of the scale. It refers to the suitability and logical consistency between the content of a measure and the measurement objective. It can also be described as an "examination" of whether the selected measurement item meets the purpose and requirements of the measurement. The review method mainly uses the expert judgment method, in which

the relevant experts evaluate the suitability of the item from theoretical and practical perspectives. The content design of this study's questionnaire is based on theory and refers to other authoritative literature. Therefore, this questionnaire has good content validity.

Sekaran and Bougie (2016) explain that validity is a test of how well the instruments, techniques or processes used to measure the concept measure the concept in question. The type of validity used in this research is construct validity which will prove how well the use is obtained in accordance with the surrounding theory designed in the test. This is assessed through convergent and discriminant validity that will be obtained from each item in the form of a question with a total score. This total score is the value obtained from the sum of all item scores. Based on the statistical measure, if it turns out that the scores of all items arranged according to the dimension of the concept of correlation with the total score, it can be said that the measuring instrument has validity. The validity of an instrument is calculated using the Kaiser–Meyer–Olkin criterion(KMO) and Bartlett test, KMO formula as follows:

$$KMO = \frac{\sum_{j \neq k} \sum r_{jk}^2}{\sum_{j \neq k} \sum r_{jk}^2 + \sum_{j \neq k} \sum p_{jk}^2}$$

where:

R = [r_{ij}] is the correlation matrix,

U = [u_{ij}] is the partial covariance matrix,

Σ = summation notation ("add up").

The Bartlett test include three parts, Chi-Square test

The Chi-Square is denoted by χ^2 . The chi-square formula is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where

O_i = observed value (actual value)

E_i = expected value.

Table 3.5
KMO and Bartlett's Test

KMO		0.953
Bartlett's Test of Sphericity	Chi-Square	7036.2
	<i>df</i>	190
	<i>p</i>	0

Source: Processed data (2022)

KMO and Bartlett tests were used for validity verification. As can be seen from Table 3.5, the KMO value was 0.953 and more significant than 0.8, indicating that the research data were very suitable for extracting information (with good validity reflected from the side).

3.5 Data Analysis Techniques

Data analysis is a systematic process to analyze the statistically collected data to see whether the resulting hypothesis has been supported by data (Sekaran & Bougie, 2016). The questionnaire is used in this study as a tool to collect data. The author will compile the questionnaire based on variables contained in this study.

In quantitative research, data analysis is carried out after the data of all respondents are collected. Data analysis activities in this study were carried out through several stages as follows:

1. Compiling data, this activity is carried out to check the completeness of the respondent's identity, complete data, and data entry following the research objectives.
2. Selecting data, this activity is carried out to check the truth of the respondents' data collected.
3. Data tabulation, the author tabulation of data with the following steps:
 - (1) Give a score on each question item
 - (2) add up the score on each question item
 - (3) compile ranking scores on each research variable
4. Analyzing data, this activity is a data processing process using the statistical formula and interpreting data to obtain a conclusion.

Likert scales will be used in this study, as techniques that is usually used for measuring opinions and attitudes (Sekaran & Bougie (2016). It measures the extent to which participants agree or disagree with a given statement, and typically range from 1 (strongly disagree) to 7 (strongly agree) with a neutral point in the middle

Table 3.6
Alternative Score

Alternative answers	Answer Range							Alternative answers	
	Very Low / Very Poor / Very Unattractive / Very Uninvasive / Very Dissatisfied / Very Unpopular	1	2	3	4	5	6		7
	Negative								positive

Source: Modifications from Sekaran and Bougie (2016)

Based on Table 3.2, it is known that the use of the Semantic Differential scale can produce answers to questions in the form of a tendency where if the respondent answers on a scale of 1-3, the respondent tends to be negative or not good. Meanwhile, if the respondent answers on a scale of 5-7, respondents tend to be positive or tend to be good. If the respondent chooses a scale of 3, the respondent answers neutrally.

SPSSAU, as a statistical web for social science, will be used to edit, code, and analyze data. Microsoft Excel version 2021 will be used to add value to the SPSSAU data produced. The data will be categorized and tabulated according to the concepts to address the purpose of the study. Result will be presented by tables, Pie charts, and bar charts.

3.5 Descriptive Data Analysis

To avoid errors caused by the abnormal value in the data, the questionnaire should be described as a descriptive statistical analysis. Measure the average number of each dimension and the standard deviation of each measurement. This article also uses Skewness and Kurtosis to measure the stability of the data results.

3.6 Multiple Linear Regression

In order to verify that there are certain influence between independent variables and dependent variables, this relationship should be tested by regression analysis

(Frost, 2020). So, this study will use multiple regression analysis. Multiple regression, also known as Multiple linear regression (MLR), is a statistical technique that uses several explanatory variables to predict the outcome of a response variable (Frost, 2020). The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables (Frost, 2020). The formula for a multiple linear regression is:

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon$$

Informations:

y = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_n = slope coefficients for each explanatory variable

ε = the model's error term (also known as the residuals)

3.7 Hypothesis Testing

Hypothesis testing, also known as statistical hypothesis testing, is a statistical inference method used to determine whether the difference between a sample and a sample or a sample and the total is caused by sampling error or essential differences. Significance testing is one of the most common hypothesis testing methods and is also one of the most basic forms of statistical inference (Sekaran & Bougie (2016). The basic principle is first to make a specific hypothesis about the characteristics of the total and then infer whether this hypothesis should be rejected or accepted through statistical inference of the sampling study (Sekaran & Bougie (2016).

To test hypothesis of this study two testing techniques will be used:

- 1) The t-test, also known as the student's t-test, has a normal distribution with an overall standard deviation σ unknown. The t-test uses the theory of t-distribution to infer the probability of a difference occurring and thus compare whether the difference between two means is significant (Ning, Z. 2015).
- 2) The F-test, most known as the joint hypotheses test, is the variance ratio test and chi-square test. It is a test under the null hypothesis (H_0) that the statistical values obey the F-distribution. It is usually used to analyze statistical models with more

than one parameter to determine whether all or some of the parameters in the model are suitable for estimating the population (Ning, Z. 2015).

3.8 Hypothesis Verification

Statistically, the hypothesis to be tested in order to decide whether to accept or reject the hypothesis can be formulated as follows:

1) H1:

H0: $\rho > 0.05$ means, there is no effect of content marketing on consumers' purchase intention.

Ha: $\rho < 0.05$ means, there is an effect content marketing on consumers' purchase intention.

2) H2:

H0: $\rho > 0.05$ means, there is no effect of product content positively on consumers' purchase intention.

Ha: $\rho < 0.05$ means, there is an effect of product content positively on consumers' purchase intention.

3) H3:

H0: $\rho > 0.05$ means, there is no effect of entertainment content positively on consumers' purchase intention.

Ha: $\rho < 0.05$ means, there is an effect of entertainment content positively on consumers' purchase intention.

4) H4:

H0: $\rho > 0.05$ means, there is no effect of interactive content positively on consumers' purchase intention.

Ha: $\rho < 0.05$ means, there is an effect of interactive content positively on consumers' purchase intention.