Development of A Curriculum Type Preference Scale for Preschool Education Vocational University

Xinyan Zhu, Masayu Dzainudin, Ying Qin Tee
Department of Early Childhood Education, Faculty of Human Development, Sultan Idris Education University, Perak, Malaysia
Corresponding Author Email: masayu@fpm.upsi.edu.my

Abstract
Exploring students' preferences for curriculum types is beneficial for further designing curriculum structures based on the differentiated needs of students, thereby improving curriculum and education quality. To make up for the lack of an effective measurement instrument for investigating the curriculum type preferences of Chinese vocational university students majoring in preschool education, a 5-point Likert scale is designed and developed in this study. The development procedure consists of three phases: (1) the scale generation phase; (2) the scale evaluation phase, including expert evaluation and target population pretesting; and (3) the scale testing phase, including first testing, second testing, and third testing. The results show that the initial scale consisted of three constructs (theory curriculum, application curriculum, and internship curriculum) and 21 items in total. By removing one item from the initial scale, 20 items in the final scale have been proven to have acceptable validity and reliability. The conclusions suggest that the curriculum type preference scale can be applied in practice, and the scale development procedure and methods in this study can serve as a theoretical reference for scale developers.
Keywords: Scale Development, Curriculum Type Preference, Preschool Education, Vocational University, 5-Point Likert Scale

Introduction
In China, it has become a trend for vocational universities to offer preschool education majors to cultivate practical early childhood education teachers (Hong, 2019). Correspondingly, a large number of students with different personality traits, learning levels, and educational needs have chosen to study preschool education in vocational universities, exhibiting significant individual differences (Hong, 2019). Curriculum is the main carrier for cultivating students in universities (Yuan, 2010; Gu & Shi, 2010), and dividing curriculum according to different standards forms curriculum types (Feng, 2007). For example, according to Zhu and Yi (2017), curriculum can be divided into social-based curriculum and child-based curriculum based on different educational philosophies; it can be divided into compulsory and optional curriculum based on different study requirements; and it can be divided into explicit

curriculum and implicit curriculum based on different curriculum expectations. In this context, understanding the curriculum type preferences of vocational university students majoring in preschool education is beneficial for further organizing curriculum types into curriculum structures in appropriate proportions according to the differentiated needs of students, thereby improving curriculum quality and better serving the cultivation of early childhood education talents (Zheng, 2018).

However, the existing curriculum structure of preschool education majors in Chinese vocational universities was too unified for all students and had not been designed based on their curriculum type preferences (Tao & Luo, 2020; Xu, 2020). In addition, theoretical research mainly investigated students' evaluation of the difficulty level of preschool education curriculum content, the rationality of curriculum quantity, the rationality of curriculum structure, and overall satisfaction with curriculum design in vocational universities, without conducting specialized research and investigation on students' preferences for curriculum types (Tao & Luo, 2020; Xu, 2020; Zheng, 2018). More importantly, in the research on curriculum surveys, interviews and qualitative analysis were mainly employed, while quantitative research only used non-scale questionnaires and descriptive statistical analysis, lacking a scale that has been rigorously validated as a measurement instrument (Tao & Luo, 2020; Wen et al., 2018; Xu, 2020).

Therefore, in order to fill the gap between ideal and reality, the purpose of this study is to develop a 5-point Likert scale of curriculum type preference for preschool education in vocational universities based on a systematic scale development procedure, as a measurement instrument to understand students' curriculum type preferences.

Methods
The scale development procedure in this study employed the three phases: scale generation, scale evaluation, and scale testing (Boateng et al., 2018; DeVellis & Thorpe, 2021; Kyriazos & Stalikas, 2018; Morgado, 2017). The scale generation phase is to develop an initial scale of the curriculum type preferences. The scale evaluation phase includes two steps: the first step is to utilize expert evaluation to judge the content validity of the scale; the second step is to utilize the target population pretesting to test the face validity of the scale. The scale testing phase includes three steps: the first step is to explore the construct validity of the scale through a first test; the second step is to verify the construct validity and measure the internal consistency of the scale through a second test; the third step is to check the test-retest reliability of the scale through a third test (Figure 1).
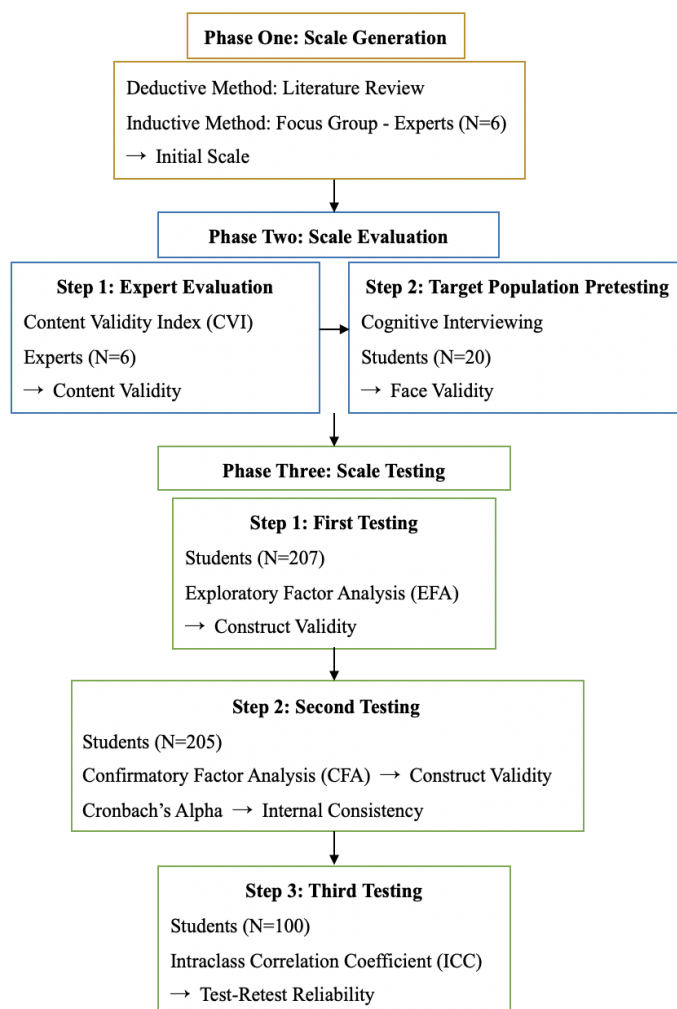
Phase One: Scale Generation

Deductive Method: Literature Review

Inductive Method: Focus Group - Experts (N=6)

→ Initial Scale

Phase Two: Scale Evaluation

Step 1: Expert Evaluation

Content Validity Index (CVI)

Experts (N=6)

→ Content Validity

Step 2: Target Population Pretesting

Cognitive Interviewing

Students (N=20)

→ Face Validity

Phase Three: Scale Testing

Step 1: First Testing

Students (N=207)

Exploratory Factor Analysis (EFA)

→ Construct Validity

Step 2: Second Testing

Students (N=205)

Confirmatory Factor Analysis (CFA) → Construct Validity

Cronbach's Alpha → Internal Consistency

Step 3: Third Testing

Students (N=100)

Intraclass Correlation Coefficient (ICC)

→ Test-Retest Reliability

Figure 1 *Scale Development Procedure*

Phase One: Scale Generation

The objective of this phase is to develop an initial scale, including defining concepts, identifying constructs, and developing corresponding items (Carpenter, 2018; Kyriazos & Stalikas, 2018). The deductive method (literature review) and the inductive method (focus groups) were employed to generate an initial scale in this study (Boateng et al., 2018; Hinkin, 2005; Morgado, 2017). First, six relevant experts were invited to participate in a focus group as recommended by Creswell and Creswell (2018) (Table 1), and a semi-structured interview protocol was developed (Appendix 1). Then, the thematic analysis method was used to analyze interview data, and the analysis procedure was as follows: (1) familiarizing data; (2) generating codes; (3) searching, reviewing, and defining themes; and (4) forming the report (Braun & Clarke, 2006).

Table 1
Expert Information

NO.	Title	Expertise	Years of Experience
Expert 1	Dean of Education Faculty of Vocational University	Vocational Education	30
Expert 2	Director of Preschool Education Department of Education Faculty of Vocational University	Preschool Education	18
Expert 3	Professor of Preschool Education	Preschool Education	40
Expert 4	Professor of Preschool Education	Preschool Education	38
Expert 5	Curriculum Expert	Curriculum & Instruction	28
Expert 6	Kindergarten Director	Preschool Education	25

Phase Two: Scale Evaluation
Step 1: Expert Evaluation
The objective of this step is to use the content validity index (CVI) (Yusoff, 2019) to measure the content validity of the scale. The procedure consists of six steps: (1) developing a content validation form (Appendix 2); (2) recruiting six experts for evaluating the relevance of items (Table 1); (3) implementing the content validity of the scale; (4) checking the domains and items of the scale; (5) providing scores for the items on the scale; and (6) calculating the CVI of the scale (Yusoff, 2019).

Step 2: Target Population Pretesting
The objective of this step is to adopt cognitive interviewing (Howard, 2018) to pretest the participants to judge the face validity (e. g., wording concerns) of the scale. 20 Chinese vocational university students majoring in preschool education were invited as participants as suggested by Lenzner et al (2016), and a semi-structured protocol of cognitive interviewing Collins (2003) was adapted in this study (Appendix 3). Moreover, a procedure of cognitive interviewing Drennan (2003) was used in this study, which includes two stages: in the concurrent interview stage, the think-aloud method is used to encourage students who are thinking about the items to verbalize their thoughts. In the retrospective interview stage, the probing technique is used to recall the memory of students who have completed the items to find existing problems. Finally, the thematic analysis method was used to analyze the data, and the analysis procedure was described in the scale generation phase.

Phase Three: Scale Testing
Step 1: First Testing
The objective of this step is to use exploratory factor analysis (EFA) (Fabrigar & Wegener, 2011) to explore the construct of the scale by conducting the first test. Specifically, principal components analysis (PCA) was used to extract factors, and varimax was used as the rotational method Taherdoost et al (2022), and statistically analyzed through the Statistical

Package for the Social Sciences (SPSS). In addition, 200 Chinese preschool education students at vocational universities were planned as respondents according to previous recommendations (Comrey & Lee, 1992; Johanson & Brooks, 2010). However, 210 questionnaires were distributed, and 207 valid questionnaires (182 females and 25 males) were recovered.

Step 2: Second Testing
The objective of this step is to confirm the construct of the scale and check the internal consistency of the scale by conducting a second test. A fresh sample with 205 Chinese vocational university students majoring in preschool education (181 females and 24 males) was tested, and the data analysis process was as follows: firstly, confirmatory factor analysis (CFA) was used to verify the construct validity of the scale (Brown, 2015). Specifically, maximum likelihood (ML) was used as an estimator (Brown, 2015), and statistically analyzed using Mplus. Subsequently, Cronbach's alpha was applied to identify the internal consistency of the scale (Tavakol & Dennick, 2011), and SPSS was used to analyze the data.

Step 3: Third Testing
The objective of this step is to measure the test-retest reliability (Guttman, 1945) of the scale by conducting a third test. This study conducted a third test in the third week after the second test and selected 100 students (88 females and 12 males) as respondents from the sample of the second test based on existing suggestions (Kennedy, 2022; Kurpius & Stafford, 2006). In addition, the intraclass correlation coefficient (ICC) was used to analyze the test-retest reliability of the scale (Koo & Li, 2016; Weir, 2005). Specifically, ICC estimates and their 95% confident intervals were calculated using SPSS based on a single-measurement, absolute-agreement, 2-way mixed-effects model (Koo & Li, 2016).

Results
Initial Scale
On the one hand, through the literature review (Bu, 2005; Yang & Cao, 2009; Yang & Kang, 2006), it was found that based on the criterion of "the proportion of theory and practice", preschool education curriculum can be divided into three types: theory curriculum, application curriculum, and internship curriculum. The theory curriculum mainly focuses on the teaching of systematic theoretical knowledge in the form of lectures, discussions, questions, analysis, and reasoning. The application curriculum is a step-by-step simulation exercise and practice based on the theoretical guidance, which contains both theoretical proportion and practical proportion. The internship curriculum is a direct experience in kindergarten. The corresponding positions of the three types can be found in the dimension of " Theoretical-Practical" (Figure 2).
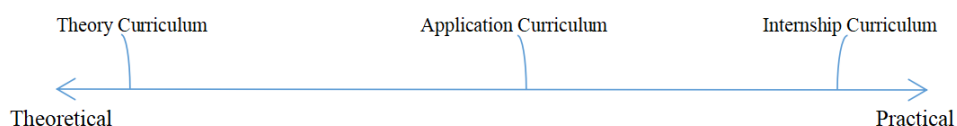

Figure 2 Dimension of "Theoretical-Practical"

On the other hand, three themes were refined through a focus group, and six experts unanimously believed that, in reality, preschool education curriculum in vocational universities can be divided into three types.

Theme 1: Theory Curriculum
Six experts unanimously agreed that courses that mainly combine logical analysis and audio-visual methods as the main teaching and learning methods, with collective teaching as the main organizational form and theoretical knowledge as the teaching content are classified as theory curriculum. *"The course objectives require students to master systematic theoretical knowledge and be able to discuss and analyze theoretical problems such as Preschool Psychology, Preschool Pedagogy, and Preschool Hygiene, which we define as theory curriculum"* (Expert 1).

Theme 2: Application Curriculum
Six experts unanimously believed that courses with both theoretical and practical elements can be classified as application curriculum, which are equipped with specialized training rooms for students to simulate and practice under the guidance of theory. *"The Montessori Teaching Method belongs to the application curriculum, as teachers need to first provide students with theoretical knowledge and then guide them to use Montessori materials in the training room"* (Expert 3).

Theme 3: Internship Curriculum
Six experts unanimously classified pure practical courses that directly go to kindergarten to observe or participate in educational activities as internship curriculum. *"The courses that require students to directly enter kindergarten to observe children's behavior, or organize children's plays and teaching, are named internship curriculum by us"* (Expert 4).

The literature review and expert discussion yielded the same finding. Therefore, this study divided the scale into three constructs: theory curriculum (TC), application curriculum (AC), and internship curriculum (IC). Afterwards, the corresponding seven items under each construct were developed, an item pool containing 21 items was generated, and a 5-point Likert initial scale was developed (Appendix 2).

Content Validity
When researchers measure the CVI, both the item-level content validity index (I-CVI) and scale-level content validity index (S-CVI) should be considered (Yusoff, 2019). According to the benchmark that the CVI estimates should be 0.83 or above with the participation of six evaluators (Polit & Beck, 2006; Polit et al., 2007), the result of the I-CVI showed that except for TC7 (I-CVI =0.5) and AC4 (I-CVI =0.67) (Table 2), the remaining items (I-CVI=1) had excellent content validity (Table 2). Other than that, the S-CVI/Ave (=0.96) and S-CVI/UA (=0.90) were higher than 0.83 (Table 2), which indicated that the content validity of the entire scale was acceptable. Afterwards, TC7 and AC4 were improved and certified based on expert suggestions.

Table 2
The Result of Content Validity

Item	I-CVI	Item	I-CVI	Item	I-CVI
TC1	1	AC1	1	IC1	1
TC2	1	AC2	1	IC2	1
TC3	1	AC3	1	IC3	1
TC4	1	AC4	.67	IC4	1
TC5	1	AC5	1	IC5	1
TC6	1	AC6	1	IC6	1
TC7	.5	AC7	1	IC7	1
S-CVI/Ave	.96	**S-CVI/UA**	.90		

Note: TC1-TC7 = Theory Curriculum Item 1-7; AC1-AC7 = Application Curriculum Item 1-7; IC1-IC7 = Internship Curriculum Item 1-7; CVI: content validity index; UA: universal agreement; I-CVI: item-level content validity index; S-CVI/Ave: scale-level content validity index based on the average method; S-CVI/UA: scale-level content validity index based on the universal agreement method (Polit & Beck, 2006; Polit et al., 2007; Yusoff, 2019).

Face Validity
The result of the pretesting showed that the target population had a clear understanding of the items and made choices, without any ambiguity, inappropriate wording, or redundant statements. Therefore, the original items were retained without any modifications.

Construct Validity - EFA
The EFA report needs to present the following three parts: (1) data inspection; (2) factor retention; and (3) factor interpretation (Howard, 2016). Firstly, the result of the data inspection showed that the value of Kaiser-Meyer-Olkin (KMO) was 0.907 (Table 3). According to the guideline for interpreting KMO values (.90 to 1.0 = marvelous; .80 to .89 = meritorious; .70 to .79 = middling; .60 to .69 = mediocre; .50 to .59 = miserable; below .50 = unacceptable) (Kaiser, 1974), this result indicated that sampling adequacy was excellent. Moreover, the result of Bartlett's test of Sphericity showed strong significance (p=0.000 <0.01) (Bartlett, 1950; Tabachnick & Fidell, 2007) (Table 3), indicating that the data was suitable for factor analysis.

Table 3
The Result of Data Inspection

Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)		.907
Bartlett's Test of Sphericity		
Approx. Chi-Square	**df**	**Sig.**
2680.240	210	.000

Note: df=degrees of freedom

Subsequently, the result of the factor retention showed that there were three factors with eigenvalues greater than one (Table 4), indicating that 21 items could be divided into three constructs (Kaiser, 1960). In addition, the cumulative percentage of variance (CPV) was 64.306% (Table 4), which met the standard of explained variance (commonly as low as 50-60% in the humanities) (Pett et al., 2003).

Table 4
The Result of Factor Retention

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.847	43.127	43.127	8.847	43.127	43.127
2	2.312	12.011	55.138	2.312	12.011	55.138
3	1.715	9.168	64.306	1.715	9.168	64.306

Finally, according to Hair et al. (2009), the values of factor loading should be at least 0.5, and above 0.7 was better. Through comprehensive consideration, this study determined that the cut-off value of the EFA factor loading was at least 0.6. The result of factor interpretation showed that except for AC1, which had a factor loading value of 0.572, all other items had values greater than 0.6 (Table 5), indicating that the correlation between the remaining 20 items and the corresponding constructs met the standard, while AC1 needs to be removed.

Table 5
The Result of Factor Interpretation

Item	Factor Loading		
	1	2	3
TC1 I like or good at *Preschool Pedagogy*	.824		
TC2 I like or good at *Preschool Psychology*	.692		
TC3 I like or good at *Preschool Hygiene*	.650		
TC4 I like or good at *Preschool Education History*	.746		
TC5 I like or good at *Introduction to Kindergarten Curriculum*	.826		
TC6 I like or good at *Professional ethics and law of Kindergarten Teachers*	.751		
TC7 I like or good at *Philosophy of Preschool Children's Education*	.775		
AC1 I like or good at *Design and Implementation of Kindergarten Education Activities*	.372	.572	
AC2 I like or good at *Teaching Method of Montessori*		.713	
AC3 I like or good at *Sensory Integration Training*		.712	
AC4 I like or good at *Storytelling of Preschool Children*		.613	
AC5 I Like or good at *Preschool Music Education (such as solfeggio, piano, Orff)*		.648	
AC6 I like or good at *Preschool Fine Art Education (such as painting, handicrafts, environmental creation)*		.608	
AC7 I like or good at *Preschool Dance Education*		.761	
IC1 I like or good at observing children's behavior as interns			.636
IC2 I like or good at interacting and communicating with children as interns			.655
IC3 I like or good at observing how kindergarten teachers organize children's activities as interns			.790
IC4 I like or good at observing how kindergarten teachers manage classes as interns			.761

IC5 I like or good at organizing children's collective teaching activities as interns	.756
IC6 I like or good at organizing children's plays as interns	.767
IC7 I like or good at organizing children's life activities as interns	.796

Note: TC1-TC7 = Theory Curriculum Item 1-7; AC1-AC7 = Application Curriculum Item 1-7; IC1-IC7 = Internship Curriculum Item 1-7.

Construct Validity - CFA

In terms of verifying the construct validity of the scale, model fit indices and standardized factor loadings are commonly used to explain the results of CFA (Jackson et al., 2009; Schreiber et al., 2006). Table 6 presents the output of Mplus for the model fit indices. According to the standard benchmark：χ^2/df (Ratio of χ^2 to df ≤2 or 3) (Cole, 1987; Schreiber et al., 2006); RMSEA (0.00 to 0.05), CFI (0.95 to 1.00), TLI (0.95 to 1.00), and SRMR (0.00 to 0.08) (Hu & Bentler, 1999; Yu, 2002), the all indices met the standards and suggested a good model fit.

Table 6
The Result of Model Fit Indices

Model Fit Index	χ^2	df	χ^2/df	RMSEA	CFI	TLI	SRMR
Estimate	539.985	210	2.57	.046	.963	.956	.072

Note: χ^2=chi-square; df=degrees of freedom; χ^2/df=ratio of chi-square to degrees of freedom; RMSEA=root mean square error of approximation; CFI=comparative fit index; TLI= Tucker-Lewis index; SRMR= standardized root mean square residual (Hu & Bentler, 1999; Schreiber et al., 2006; Yu, 2002).

Figure 3 presents the output of Mplus for the standardized factor loadings. Consistent with the factor loading standard of EFA, this study determined that the cut-off value of the CFA factor loading was at least 0.6. The result showed that the factor loading values of all items were greater than 0.6, indicating that the correlation between all factors and the corresponding items was within the acceptable range.
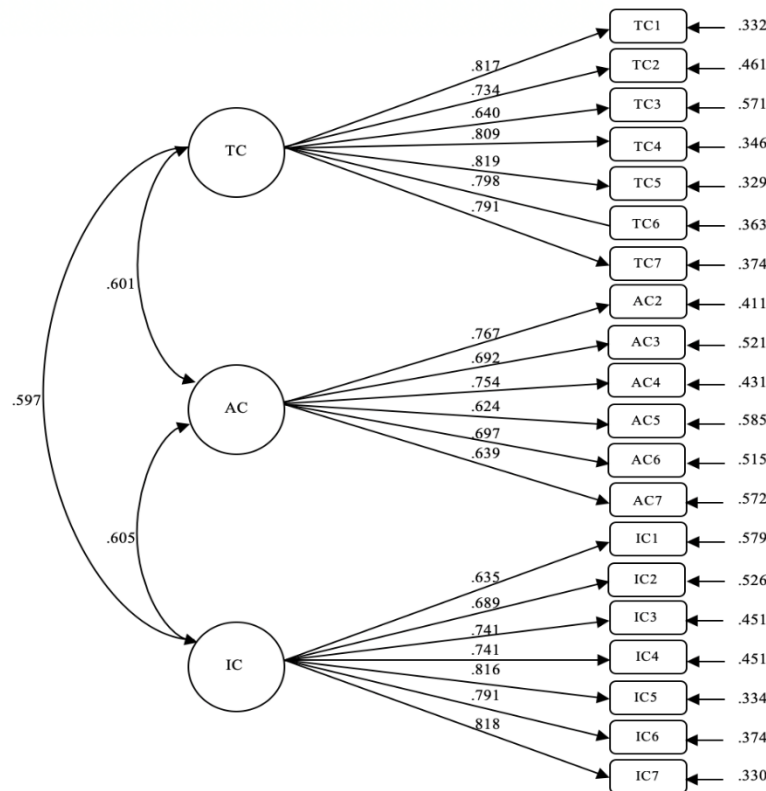
Figure 3 *The Standardized Factor Loadings*

Internal Consistency

The result of the internal consistency showed that the values of Cronbach's alpha of the whole scale and three subscales were between 0.851 and 0.929 (Table 7). Referring to the interpretation for Cronbach's alpha (≥.9 = Excellent; ≥.8 = Good; ≥.7 = Acceptable; ≥.6 = Questionable; ≥.5 = Poor; ≤.5 = Unacceptable) (George & Mallery, 2003), the result suggested that both the whole scale and subscales had at least good reliability.

Table 7
The Result of Internal Consistency

	Whole Scale	TC	AC	IC
Cronbach's Alpha	.929	.912	.851	.899

Test-Retest Reliability

The result showed that the ICC estimate of all items ranged from 0.833 to 0.977, so the estimate of ICC with the 95% confidence interval of all items ranged from 0.762 to 0.985 (Table 8). Relying on ICC Interpretation (ICC < 0.5 = Poor; 0.5 ≤ ICC < 0.75 = Moderate; 0.75 ≤ ICC < 0.9 = Good; 0.9 ≤ ICC = Excellent) (Koo & Li, 2016; Portney & Watkins, 2000), the results showed that all items at least achieved good reliability.

Table 8
The Result of Test-Retest Reliability

Item	ICC (95% CI)		Item	ICC (95% CI)		Item	ICC (95% CI)	
TC1	.917	(.878–.943)	-	-		IC1	.977	(.966-.985)
TC2	.967	(.951-.977)	AC2	.893	(.844-.926)	IC2	.901	(.856-.933)
TC3	.833	(.762-.885)	AC3	.934	(.903-.955)	IC3	.897	(.851-.929)
TC4	.839	(.768-.889)	AC4	.834	(.764-.885)	IC4	.884	(.832-.920)
TC5	.954	(.932-.969)	AC5	.867	(.808-.909)	IC5	.883	(.832-.920)
TC6	.841	(.772-.890)	AC6	.880	(.827-.918)	IC6	.920	(.884-.946)
TC7	.918	(.879-.944)	AC7	.915	(.876-.942)	IC7	.956	(.935-.970)

Note: TC1-TC7 = Theory Curriculum Item 1-7; AC1-AC7 = Application Curriculum Item 1-7; IC1-IC7 = Internship Curriculum Item 1-7; ICC = Intraclass Correlation Coefficient; 95% CI =95% Confidence Interval.

Discussion
The findings from this study showed that by improving TC7, AC4, and removing AC1, a scale containing 20 items was developed and can be used in a specific context (Appendix 4). During the development process, validity and reliability were analyzed, and the following discussion focuses on these two sections:

Validity
A test is valid if it measures what it claims to measure (Kelley, 1927). This study tested the content validity, face validity, and construct validity of the scale. Firstly, the results of the content validity showed that the correlation between TC7 (I-CVI=0.5) and AC4 (I-CVI=0.67) with their corresponding domains still needs improvement (Table 2). For the original statement of TC7 (Preschool Family Education), experts believed that some universities do not conduct this course purely as a theory curriculum, but rather incorporate a certain proportion of practice, such as designing situational performances. Therefore, experts suggested replacing this item with "Philosophy of Preschool Children's Education", which is more relevant to the domain of theory curriculum. For the original statement of AC4 (Preschool Child Care), experts believed that some universities integrated this course into "Preschool Hygiene" when carrying out this course. Therefore, experts suggested replacing this item with "Storytelling of Preschool Children", which is more relevant to the domain of application curriculum.

Moreover, the pretest on the target population found that the face validity of the scale was good, and respondents were able to clearly understand and make choices about the statement of the item. The reason is likely because the statement of all items is the familiar

course name for the students, without ambiguity, inappropriate wording, or redundant statements.

Finally, from the EFA result, it can be clearly seen that the factor loading estimate of AC1 (0.572) was relatively low in the domain of the corresponding application curriculum, and also has a weak correlation in the theory curriculum (0.372) (Table 5). The possible reason for this is that the original item statement of AC1 is "Implementation of Kindergarten Education Activities", and this course content mainly designs kindergarten education activities and conducts corresponding simulation teaching based on theoretical knowledge, which belongs to the application curriculum. However, in reality, it has been observed that some lecturers who teach this course lack practical experience in kindergarten and the practical ability to guide students, leading to a large extent to teaching this course as a purely theoretical course, resulting in differences in students' positioning of this course. In addition, it was due to the strict screening of items using EFA (factor loading of 0.6 or above) and the removal of poor-quality items that CFA verified good construct validity in the remaining 20 items.

Reliability
Bruton et al (2000) stated that reliability refers to the repeatability or consistency of a measuring test or quantitative research. This study considered two commonly measured reliabilities: internal consistency and test-retest reliability. The result showed that the whole scale and subscales had good to excellent internal consistency, and the test-retest reliability of all items also achieved good to excellent stability. The reason for this is that on the one hand, it reflects that the scale indeed has good reliability, and on the other hand, it can also reflect that students' preferences for curriculum types and the degree of preference are relatively stable.

Conclusion
The findings of this study showed a initial curriculum type preference scale for preschool education in vocational universities, consisting of three constructs (theory curriculum, application curriculum, and internship curriculum) and a total of 21 items. Moreover, the initial scale was tested for validity and reliability, and a final scale consisting of 20 items was developed and determined.

This study has both practical and theoretical implications. A curriculum type preference scale for preschool education in vocational universities was systematically developed for the first time, which can be used as an instrument to measure students' preferences for curriculum types in practice. In addition, the development process and methods used in this study can provide theoretical support for scale developers.

However, scale development is a long-term and complex process, and inevitably has limitations. This study used inductive methods during the scale generation phase to understand the actual status of curriculum design and serve as a reference for determining construct and developing items. However, it can hardly provide a more comprehensive understanding of the actual situation of curriculum design only through focus groups. For example, in AC1 (Implementation of Kindergarten Education Activities), due to the failure to comprehensively investigate the diverse understanding of the course type among students in reality and the failure to remove the item, the correlation between the item and the

corresponding construct was weak. It can be said that there is a certain limitation in the research method used in this study.

Therefore, this study suggests that in future research, the inductive methods in the item generation phase can specifically use the questionnaire survey method, with the aim of expanding the survey scope and providing more comprehensive reference information for developing items.

References

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3*(2), 77-85.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R. & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in Public Health, (6),* Article149. https://doi: 10.3389/fpubh.2018.00149

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Resreach in Psychology, 3*(2), 77–101.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, second edition.* Guilford Press.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: what is it, and how is it measured? *Physiotherapy, 86*(2), 94-99. https://doi.org/10.1016/S0031-9406(05)61211-4

Bu, S. M. (2005). On objective orientation and curriculum design for undergraduate preschool education program. *Teacher Education Research, 17*(03), 20-24. https://doi:10.13445/j.cnki.t.e.r.2005.03.003

Carpenter, S. (2018). Ten steps in scale development and reporting: a guide for researchers. *Communication Methods and Measures, 12*(1), 25-44, https://doi:10.1080/19312458.2017.1396583

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*(4), 584–594. https://doi.org/10.1037/0022-006X.55.4.584

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research,* 12, 229-238.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd edition).* Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches.* SAGE Publications.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: theory and applications (fifth edition).* SAGE Publications.

Drennan, J. (2003). Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57–63.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory Factor Analysis.* Oxford University Press.

Feng, J. J. (2007). *Foundation of modern pedagogy (3nd ed.).* Nanjing Normal University Press.

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference. 11.0 Update (4th edn.).* Allyn & Bacon.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*(4), 255–282.

Gu, M. Y., & Shi, Z. Y. (2010). *Interpretation of the national medium and long term education reform and development planning outline (2010-2020).* Beijing Normal University Press.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Multivariate data analysis. (7th, Ed.).* Prentice Hall.

Hinkin, T. R. (2005). *Scale development principles and practices.* In R. A. Swanson & E. F. Holton (Eds.), Research in organizations: Foundations and methods in inquiry (pp. 161-180). Berrett-Koehler Publishers.

Hong, X. M. (2019). *China education reform and openning-up 40 years.* Beijing Normal University Press.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *International Journal of Human-Computer Interaction, 32*(1), 51-62. DOI: 10.1080/10447318.2015.1087664

Howard, M. C. (2018). Scale pretesting. *Practical Assessment, Research, and Evaluation, 23*(5),1-14. https://doi.org/10.7275/hwpz-jx61

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1–55. https://doi.org/10.1080/ 10705519909540118

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6–23. https://doi.org/10.1037/a0014694

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and psychological, 70*(3), 394-400. https://doi.org/10.1177/0013164409355692

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1),141-151. https://doi.org/10.1177/001316446002000116

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika,* 39, 31-36. https://doi:10.1007/BF02291575

Kelley, T. L. (1927). *Interpretation of educational measurements.* Macmillan.

Kennedy, I. (2022). Sample size determination in test-retest and cronbach alpha reliability estimates. *British Journal of Contemporary Education 2*(1), 17-29. DOI: 10.52589/BJCE-FY266HK9

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kurpius, S. E. R., & Stafford, M. E. (2006). *Testing and measurement: A user-friendly guide.* Sage Publications.

Kyriazos, T. A., & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardization process. *Psychology,* 9, 2531-2560. https://doi.org/10.4236/psych.2018.911145

Lenzner, T., Neuert, C., & Otto, W. (2016). Cognitive Pretesting (Version 2.0). (GESIS Survey Guidelines).GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_010

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*(5), 1–20. https://doi.org/10.1186/s41155-016-0057-1

Pett, M. A., Lackey, N. R., & Sullivanet, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Sage Publications Inc.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467. https://doi:10.1002/nur.20199

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice. 2nd edition.* Prentice Hall Health, Upper Saddle River.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338. http://dx.doi.org/10.3200/JOER.99.6.323-338

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics. Fifth Edition.* Pearson Education Inc.

Taherdoost, H., Sahibuddin, S., & Jalaliyoon, N. (2022). Exploratory factor analysis; concepts and theory. *Advances in Applied and Pure Mathematics, 27,* 375-382.

Tao, C. P., & Luo, Y. (2020). Some thoughts on the curriculum design of preschool education specialty in higher vocational education ——based on the investigation and research on the satisfaction level of preschool education curriculum. *Education Teaching Forum,* (19), 382-384.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's Alpha. *International journal of Medical Education,* 2, 53-55. https://doi: 10.5116/ijme.4dfb.8dfd

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research, 19*(1), 231–240.

Wen, P., Huang, R. J., & She, Y. B. (2018). An investigation and analysis on the curriculum of preschool education in higher vocational education—from the perspective of selective education. *Journal of Educational Development,* (12), 80-84. https://doi:10.16215/j.cnki.cn44-1371/g4.2018.12.016

Xu, L. J. (2020). Current situation of majoring in pre-school education in vacational college from the student perspective. *Journal of Bengbu University, 9*(3), 73-78. https://doi:10.13900/j.cnki.jbc.2020.03.017

Yang, L. J., & Cao, Li. (2009). The reconstruction of personnel cultivation mode of undergraduate preschool education program in higher normal colleges and universities. *Journal of Educational Science of Hunan Normal University, 8*(04), 82-85.

Yang, L. J., & Kang, D. (2006). Thoughts on several problems in the training of undergraduate students majoring in preschool education in normal universities. *Teacher Education Research,* (05), 70-74.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Unpublished doctoral dissertation). University of California Los Angeles, Los Angeles, CA.

Yuan, Z. G. (2010). *Contemporary pedagogy.* Educational Science Press.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zheng, M. (2018). An investigation report on the current situation of curriculum provision of preschool education major in Higher Vocational Colleges—a case study of a higher

vocational college in Hebei Province. *Journal of Hebei Youth Administrative Cadres College,30*(03),55-59. https://doi:10.16278/j.cnki.cn131260/d.2018.03.011

Zhu, D. Q., & Yi, L. Y. (2017). *Introduction to pedagogy. (3nd ed.).* Southwest Normal University Press.

Appendix 1: Interview Protocol for Initial Scale Development

1.In curriculum design, which criterion is more applicable or common for categorizing preschool education curriculum types?

2. According to above criterion, what types of preschool education curriculum can be divided into?

3. What specific courses are available for each curriculum type?

Appendix 2: Content Validation Form

Dear Experts,

This scale contains 3 domains and 21 items related to preference for preschool education curriculum types in vocational universities. We need your expert judgement on the degree of relevant of each item to the measured domains. Your review should be based on the definition and relevant terminologies that are provided to you. Please be as objective and constructive as possible in your review and use the following rating scale:

Degree of Relevance

1=the item is not relevant to the measured domain

2=the item is somewhat relevant to the measured domain

3=the item is quite relevant to the measured domain

4=the item is highly relevant to the measured domain

Domain 1：Theory Curriculum (TC)	
Definition：Theory curriculum mainly focuses on the teaching of systematic theoretical knowledge in the form of lectures, discussions, questions, analysis and reasoning.	
TESTED ITEMS	**RELEVANCE**
TC1. I like or good at *Preschool Pedagogy*	1 2 3 4
TC2. I like or good at *Preschool Psychology*	1 2 3 4
TC3. I like or good at *Preschool Hygiene*	1 2 3 4
TC4. I like or good at *Preschool Education History*	1 2 3 4
TC5. I like or good at *Introduction to Kindergarten Curriculum*	1 2 3 4
TC6. I like or good at *Professional ethics and law of Kindergarten Teachers*	1 2 3 4
TC7. I like or good at *Preschool Family Education*	1 2 3 4
Domain 2：Application Curriculum (AC)	
Definition：Application curriculum is a step-by-step simulation exercise and practice based on the theoretical guidance. It contains both theoretical proportion and practical proportion.	
TESTED ITEMS	**RELEVANCE**
AC1. I like or good at *Design and Implementation of Kindergarten Education Activities*	1 2 3 4
AC2. I like or good at *Teaching Method of Montessori*	1 2 3 4
AC3. I like or good at *Sensory Integration Training*	1 2 3 4

AC4. I like or good at *Preschool Child Care*	1 2 3 4
AC5. I Like or good at *Preschool Music Education (such as solfeggio, piano, Orff)*	1 2 3 4
AC6. I like or good at *Preschool Fine Art Education (such as painting, handicrafts, environmental creation)*	1 2 3 4
AC7. I like or good at *Preschool Dance Education*	1 2 3 4

Domain 3：Internship Curriculum (IC)
Definition： Internship curriculum is a direct experience in the kindergarten.

TESTED ITEMS	RELEVANCE
IC1. I like or good at observing children's behavior as interns	1 2 3 4
IC2. I like or good at interacting and communicating with children as interns	1 2 3 4
IC3. I like or good at observing how kindergarten teachers organize children's activities as interns	1 2 3 4
IC4. I like or good at observing how kindergarten teachers manage classes as interns	1 2 3 4
IC5. I like or good at organizing children's collective teaching activities as interns	1 2 3 4
IC6. I like or good at organizing children's plays as interns	1 2 3 4
IC7. I like or good at organizing children's life activities as interns	1 2 3 4

Appendix 3：Protocol of Cognitive Interviewing
In the concurrent interview stage: think-aloud method
How did you feel about answering this item?
I noticed you hesitated before you answered – what were you thinking about?
How easy or difficult did you find this item to answer? Why do you say that?
In the retrospective interview stage: probing techniques
Please recall, which items were difficult for you to understand?
Why are these items difficult for you to understand?
How would you improve these difficult to understand items?

Appendix 4：Final Scale

No.	Curriculum Type Preferences	Strongly Agree (5)	Agree (4)	Neutral (3)	Disagree (2)	Strongly Disagree (1)
TC	**Theory Curriculum**					
TC1	I like or good at *Preschool Pedagogy*					
TC2	I like or good at *Preschool Psychology*					
TC3	I like or good at *Preschool Hygiene*					
TC4	I like or good at *Preschool Education History*					
TC5	I like or good at *Introduction to Kindergarten Curriculum*					

TC6	I like or good at *Professional ethics and law of Kindergarten Teachers*					
TC7	I like or good at *Philosophy of Preschool Children's Education*					
AC	**Application Curriculum**					
AC2	I like or good at *Teaching Method of Montessori*					
AC3	I like or good at *Sensory Integration Training*					
AC4	I like or good at *Storytelling of Preschool Children*					
AC5	I Like or good at *Preschool Music Education (such as solfeggio, piano, Orff.)*					
AC6	I like or good at *Preschool Fine Art Education (such as painting, handicrafts, environmental creation)*					
AC7	I like or good at *Preschool Dance Education*					
IC	**Internship Curriculum**					
IC1	I like or good at observing children's behavior as interns					
IC2	I like or good at interacting and communicating with children as interns					
IC3	I like or good at observing how kindergarten teachers organize children's activities as interns					
IC4	I like or good at observing how kindergarten teachers manage classes as interns					
IC5	I like or good at organizing children's collective teaching activities as interns					
IC6	I like or good at organizing children's plays as interns					
IC7	I like or good at organizing children's life activities as interns					