

Title: Identify the typology of Early Childhood education through Parents' perception: A Latent class analysis of ECPP:2016

Abstract: A child's early years are the foundation for his or her future development, providing a strong base for lifelong learning and learning abilities, including cognitive and social development. Yet, we know little early childhood education preschool, since to date the majority of studies on early childhood focus on children's participation in and barriers to participating in nonparental care arrangements. The purpose of this study is to examine the different typologies of early childhood education from a large nationally representative dataset (ECPP:2016) using latent class analysis (LCA). We found five significantly different typologies (i.e., latent class) based on parents' perception. Based on these three typologies, we discuss implications for future early childhood education research and practice.

Background: Educators and policymakers are pressed to respond to the current crisis in American public education. These deficiencies are most severe in large, low-income urban settings, where children are living in neighborhoods that are characterized by a disproportionate number of family and community risk factors and overburdened school and public service agencies (Children's Defense Fund, 1997; U.S. Department of Education, 1996). Concerns about the crisis in public education have led to the establishment of the National Education Goals (National Education Goals Pane, 1999). The goals represent a strategic plan to enhance learning opportunities for all students by targeting what experts believe to be the most fundamental components of effective intervention. One of the cardinal foci of this strategic national plan is school readiness. This goal and its accompanying objectives highlight the need for quality early childhood educational programs and emphasize the importance of establishing and maintaining substantial parent involvement to promote student learning. However, there still limit information from the empirical typology analysis about the early childhood education.

Research Question: There are three research questions that we focus on this study: (1) Does there exist latent classes for the believe of early childhood education from parents? (2) How children and family background related to the estimated latent classes? and (3) Do latent classes effect children's learning?

Data Sources: This study is a secondary analysis of Early Childhood Program Participation Survey (ECPP:2016). ECPP:2016 is part of the National Household Education Surveys Program (NHES:2016), which is the flagship household survey program of the National Center for Education Statistics (NCES). NHES:2016 used a nationally representative address-based sample covering the 50 states and the District of Columbia. ECPP:2016 focus on young children's care and education before kindergarten. The surveys typically focus on children's participation in and barriers to participating in nonparental care arrangements; what activities the family does with the child, such as reading, singing, and arts and crafts; and what the child is learning, such as counting, recognizing the letters of the alphabet, and reading. Parents are the respondents.

The whole survey in this study children in the United States from birth through the age of 5 who were not yet enrolled in kindergarten by 2016. After removing the data set which miss all the value in the indicator, $n = 3871$ children are included in this study.

Methods: We used the three-step LCA model under the framework of mixture model (Jung & Wickrama, 2008; Nylund, Asparouhov, & Muthén, 2000 in this study. MPLUS (Muthén & Muthén, 2007) was used to conduct the latent class analysis (see appendix for MPLUS code). *Figure 1* details the components of the LCA model. Latent class is first estimated based on 10 indicators about the perception from parents about the main reason for choosing care (i.e., what factors were important to parents when choosing a care arrangement). *Table 1* shows the descriptive statistics of the indicators before transformation. All indicators (originally measured in four level Likert) are dichotomized by setting ‘not at all important’ and ‘a little important’ as 0 and ‘somewhat important’ and ‘very important’ as 1, when input into the LCA model. Five-class model is accepted based on the model fit statistics in general (see *Table 2*).

Then, six covariate variables are used to fit five-class model. These covariates are related to the background information of the children and their family (see *Table 4* for the descriptive analysis of the covariates in original scale). Again, we dichotomized all covariates with the threshold as the middle measurement level in the original scale. In particular, the variable about disability is based on 13 specific disability or disturbance items. If the child has at least one condition, he/she will have value of 1. Otherwise, it is transformed as 0 (i.e., do not have any kinds of disability or disturbance).

Finally, to assess the interpretation of each of the five typologies as well as to explore the reasons that students have different learning performance at early age. Four learning outcome variables are used as the dependent variable (see *Table 6* for descriptive analysis of the distal outcome in original scale) in the third step with estimated latent class as independent variable. All outcome variables are also dichotomized for interpretation purpose. We need to note that children under the age of 2 (884 observations in total) are not required to answer the questions about children’s learning. Consequently, all missing value because of age will be set as 0, which correspond to the situation that the children do not master these skills.

Results: In general, LCA with five latent class have the best model fit. Based on the model fit statistics, five-class result has the lowest information criteria values ($AIC = 30048.916$ and $BIC = 30215.438$) with the highest entropy value of 0.681. Except for the first latent class, probabilities all exceeding 0.8 for the most likely membership in each class indicating a good model fit and separation of individuals across the three latent class model (see *Table 3*). These information shows that there is a limitation in the overall model fit. However, five-class model has the best performance under the framework of LCA.

Thus, based on the responses to the survey items, we named the three dropout typologies as latent class 1, latent class 2, latent class 3, latent class 4, and latent class 5. These typologies represented 7.31%, 48%, 36.4%, 1.42%, and 6.82% of the sample respectively. Consequently, latent class 2 represent most common response pattern, followed by latent class 3. In general, ‘reliability’ and ‘learning activity’ are the two most important factors with high agreement. By contrast, ‘rating on the website’ and ‘religious orientation’ have lowest importance with low agreement (see *Table 1*).

Figure 2 details the response patterns of each of the five identified typologies to the 10 dichotomously scaled indicators. The x-axis provides each of the items, while the y-axis presents the proportion of each group that responded “important” or “not important. As demonstrated in *Figure 2*, parents in latent class 1 pay much more attention to ‘the learning activity’, ‘spending time with other kids his/her age’, ‘number of other children’ rather than ‘location’, ‘cost’, and ‘recommendation from friend and family’. They also give a relatively higher importance to the

‘religious orientation’. Parents in the Latent class 2 (as the most typical group) gives relative high importance to all indicators compared with other latent classes. In contrast, the latent class 4 gives relative low importance to almost all indicators compared with other latent classes, except ‘spending time with other kids his/her age’ and ‘recommendation from friend and family’. Latent class 3 and latent class 4 both give relative high importance value to the ‘location’, ‘cost’, ‘reliability’, and ‘recommendation from friend and family’. However, latent class 5 gives relative low importance to ‘the learning activity’, ‘spending time with other kids his/her age’, and ‘number of other children’ rather than ‘location’ and high importance to ‘caregiver is able to provide care’.

Based on the result from second-step of LCA model, we take latent class 2 as reference group since it is the most typical class in the sample (see *Table 5*). Children in latent class 1 are significantly more likely to come from a high income and white family, which may partially explain why their parents give low importance value on ‘location’, ‘cost’, and ‘reliability’. Children in latent class 3 are significantly more likely to come from a small and white family. Children in latent 4 is significantly more likely to be disable and comes from a big family, which may partially explain why their parents give low importance value on almost all indicators. Finally, children from latent class 5 is significantly more likely to come from a poor and white family, which may partially explain why their parents give high importance value on ‘location’, ‘cost’, and ‘reliability’.

Based on the third-step of LCA model (see *Table 7*), the children from latent class 1 have best outcome on all variables and children from latent class 5 have worst outcome on all variables compared with other latent classes. Looking back to response pattern in indicators for these two latent class, we could see the importance of ‘learning activity’ and the interaction with ‘other children’ in preschool education. Similarly, children from the latent class 3 also have a relative better performance on almost all variable, while their parents also pay much more attention to ‘learning activity’ and the interaction with ‘other children’. Another explanation is that children from latent class 1 is more likely to comes from the rich family, who are willing to find the good preschool education program or other types of child care regardless of ‘cost’, ‘location’, and ‘reliability’. Based on the evidence from Pearson pairwise test (see *Table 8*), children from latent class 2 and latent class 4 have no significant difference on all distal outcomes. Surprisingly, their parents’ perceptions are very different with parents in latent class 1 give high value on all indicators while parents in latent class 4 give low values. This partially indicates that it is hard to make really contributing decision when parents give high importance to all factors in reality.

Conclusion: These findings are novel and significant extensions of the past early childhood education research for three main reasons. First, our findings give a five distinctive latent class based on parents’ perception about early childhood education. Second, the background information of the kids and family (except gender), to some extent, can explain the difference among the latent classes. Finally, the latent class has significant difference in their learning outcome, which gives some insight to the early childhood education practice. However, this study also has some limitations. For example, the overall model fit in LCA need to be improved and the accuracy of most likely probability in first latent class is low. To fixed this, we may need to incorporate more indicators in LCA from other domains in the survey. Meanwhile, the results from the three-step LCA is not correlational rather than casual. The longitudinal data analysis using similar LCA methods can be explore in the future research.

Reference

- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302-317.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569.
- Muthén, L. K., & Muthén, B. O. (2007). *Mplus: Statistical analysis with latent variables user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Children's Defense Fund. (1997). *Children's Defense Fund Annual Report: Celebrating 25 Years*. Children's Defense Fund, Washington, DC.
- Thomas M., Smith B., Aronstamm Y., Susan P. C., Marianne P., Nabeel A., Mary R. R., & Yupin B., (1996). *The Condition of Education 1996*. U.S. Department of Education Office of Educational Research and Improvement, Washington, DC.
- National Education Goals Pane, (1999). *The National Education Goals Report: Building A Nation of Learners 1999*, Washington, DC.

Appendix

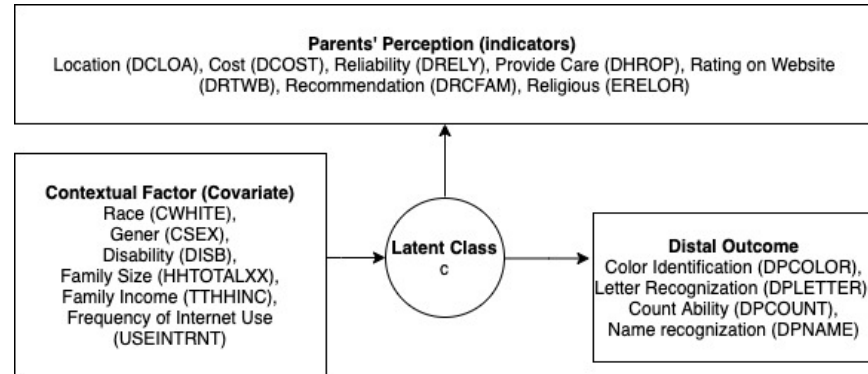


Figure 1: Latent Class Analysis (LCA) model for early childhood education

Table 1: Descriptive Analysis of the Indicator variables of LCA

Variables	N	Min	Max	Mean	SD	ECPP:2016
Location of arrangement	3871	1	4	3.434	0.790	CDLOA
Cost of arrangement	3871	1	4	3.247	0.903	DCOST
Reliability of arrangement	3871	1	4	<u>3.813</u>	<u>0.516</u>	DRELY
The learning activity at the arrangement	3871	1	4	<u>3.719</u>	<u>0.563</u>	DLERN
The child spending time with other kids his/her age	3871	1	4	3.550	0.773	DCHIL
The times during the day that this caregiver is able to provide care	3871	1	4	3.560	0.752	DHROP
The number of other children in the child's care group	3871	1	4	3.106	0.893	DNBGRP
Ratings on a website	3871	1	4	<u>2.576</u>	<u>1.135</u>	DRTWEB
Recommendation from friends and family	3871	1	4	3.192	0.967	DRECFAM
Religious orientation of the program	3871	1	4	<u>2.022</u>	<u>1.136</u>	DRELOR

Table 2: LCA results and fit statistics for Early Childhood Education

Model	AIC	Adjusted BIC	-Log likelihood	- BLRT	P-value	LMR	P-value	Entropy
Two Classes	30756.580	30821.339	15357.290	16199.000	0.000	1665.097	0.000	0.623
Three Classes	30456.748	30555.428	15196.374	15357.290	0.000	318.329	0.009	0.671
Four Classes	30202.608	30335.208	15058.304	15196.374	0.000	273.135	0.000	0.663
Five Classes	30048.916	30215.438	14970.458	15058.304	0.000	173.779	0.000	0.681
Six Classes	30022.038	30222.481	14946.019	14970.458	0.000	48.346	0.017	0.647

Table 3: Classification Probabilities for the Most Likely Latent Class Membership

Latent Classes	1	2	3	4	5
1	0.531	0.220	0.236	0.006	0.008
2	0.011	0.843	0.142	0.000	0.004
3	0.033	0.132	0.807	0.002	0.026
4	0.051	0.000	0.076	0.816	0.057
5	0.005	0.030	0.111	0.009	0.845

**Figure 2:** Latent class pattern**Table 4:** Descriptive Analysis of the Covariates

Variables	N	Min	Max	Mean	SD	ECPP:2016
Children's Race (white or not)	3871	0	1	0.806	0.396	CWHITE
Children's Gender (male or not)	3871	0	1	0.514	0.499	CSEX
Whether Children Has any Kind of Disability and Disturbance	3871	0	1	0.123	0.327	HDINTDIS, HDSPEECHX, HDDISTRBX, HDDEAFIMX, HDBLINDX, HDORTHOX, HDAUTISMX, HDPDDX, HDADDX, HDLEARNX, HDDELAYX, HDTRBRAIN, HDOTHERX
Number of people living in the family	3871	0	9	3.813	1.188	HHTOTALXX
Frequency of using Internet	3871	0	4	2.719	0.489	USEINTRNT
Total income of the family	3871	0	9	2.550	2.802	TTLHHINC

Table 5: Mean and Odd Ratio for Covariates across Latent Class with Latent Class 2 as Reference Group

Covariates	Latent class 1 (7.31%)			Latent class 3 (36.4%)			Latent class 4 (1.42%)			Latent class 5 (6.82%)		
	mean	Odd Ratio	P-value	mean	Odd Ratio	P-value	mean	Odd Ratio	P-value	mean	Odd Ratio	P-value
Disability	0.131	1.140	0.641	0.184	1.202	0.290	<u>1.081</u>	<u>2.948</u>	<u>0.005**</u>	-0.554	0.574	0.119
Number of People	0.201	1.223	0.292	<u>-0.559</u>	<u>0.572</u>	<u>0.000***</u>	<u>0.879</u>	<u>2.408</u>	<u>0.004**</u>	0.167	1.182	0.335
Internet Use	-0.255	0.775	0.683	0.801	2.229	0.214	-0.060	0.942	0.942	-0.028	0.972	0.969
Family Income	<u>0.426</u>	<u>1.530</u>	<u>0.055</u>	0.053	1.054	0.678	-0.371	0.690	0.241	<u>-1.094</u>	<u>2.985</u>	<u>0.000***</u>
Gender	0.096	1.101	0.598	0.077	1.081	0.499	-0.215	0.806	0.519	0.132	1.141	0.427
Race	<u>1.479</u>	<u>4.390</u>	<u>0.000***</u>	<u>1.008</u>	<u>2.739</u>	<u>0.000***</u>	0.269	1.309	0.428	<u>0.732</u>	<u>2.080</u>	<u>0.002**</u>

Table 6: Descriptive Analysis of the Distal Outcome

Variables	N	Min	Max	Mean	SD	ECPP:2016
Color identification	3871	0	2	1.373	0.863	DPCOLOR
Letter reorganization	3871	0	3	1.435	1.214	DPLETTER
Count Ability	3871	0	5	2.274	1.635	DPCOUNT
Whether kid can write the first word even backward	3871	0	1	0.385	0.486	DPNAME

Table 7: Mean and Standard Deviation of Outcome across Latent Classes

Outcome Variables	Latent Class 1		Latent Class 2		Latent Class 3		Latent Class 4		Latent Class 5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Color identification	<u>0.903</u>	0.027	0.776	0.013	0.865	0.016	0.715	0.075	<u>0.654</u>	0.053
Letter reorganization	<u>0.722</u>	0.039	0.588	0.016	0.596	0.022	0.589	0.083	<u>0.385</u>	0.058
Count Ability	<u>0.810</u>	0.036	0.664	0.015	0.674	0.021	0.527	0.084	<u>0.514</u>	0.058
Whether kid can write the first word even backward	<u>0.52</u>	0.054	0.516	0.016	0.514	0.022	0.452	0.084	<u>0.307</u>	0.041

Table 8: Pearson Pairwise Test of Mean with Latent Class 2 as Reference Group

Outcome Variables	Latent Class 1		Latent Class 3		Latent Class 4		Latent Class 5	
	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value
Color identification	<u>16.195</u>	<u>0.000***</u>	<u>14.470</u>	<u>0.000***</u>	0.645	0.422	4.988	0.026
Letter reorganization	<u>9.084</u>	<u>0.003**</u>	0.074	0.786	0.000	0.998	<u>11.449</u>	<u>0.001***</u>
Count Ability	<u>13.000</u>	<u>0.000***</u>	0.113	0.737	2.556	0.110	6.305	0.012
Whether kid can write the first word even backward	<u>20.625</u>	<u>0.000***</u>	0.005	0.945	0.569	0.451	<u>17.204</u>	<u>0.000***</u>

MPLUS Code

```
Title: Three-step LCA Model for Early Childhood Education
Data: File = C:\Users\ATS-Workshop\Desktop\data.dat ;
Variable: NAMES = dcloa dcost drely dlern dchil dhrop dnbgrp drtweb
             drecfam drelor rcnow ncnow cpnnowx DISB hhtotalxx
             useintrnt ttlhhinc csex cwhite
             dpcolor dpletter dpcount dpname ;
MISSING = ALL(999) ;
USEVARIABLES = dcloa dcost drely dlern dchil dhrop
              dnbgrp drtweb drecfam drelor ;
CATEGORICAL = dcloa dcost drely dlern dchil dhrop
              dnbgrp drtweb drecfam drelor ;
CLASS = c(5) ;
AUXILIARY = (R3STEP) DISB hhtotalxx
            useintrnt ttlhhinc csex cwhite;
! Command Syntax for step 2
AUXILIARY = dpcolor (BCH) dpletter (BCH)
            dpcount (BCH) dpname (BCH) ;
! Command Syntax for step 3
Analysis: TYPE = MIXTURE ;
Output: tech10 ;
Savedata: SAVE = CPROBABILITIES ;
          FORMAT=FREE;
          FILE = C:\Users\ATS-Workshop\Desktop\CPROBS-011.DAT ;
          ESTIMATES = C:\Users\ATS-Workshop\Desktop\MIXEST-011.DAT ;
```