International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3487–3496

Latent Profile Analysis (LPA) of Motivated Strategies for Learning Questionnaire (MSLQ) in the Indian Context

Rajib Chakraborty[1], Dr. Vijay Kumar Chechi[2]

*[1]Assistant Professor and Research Scholar, School of Education, Lovely Professional University, Phagwara, Punjab, India, Email: rajib.22752@lpu.co.in**
[2]HOD and Professor, School of Education, Lovely Professional University, Phagwara, Punjab, India, Email: vijay.chechi@lpu.co.in

Abstract

Latent profile analysis (LPA) was conducted on five sub-scales of the famous Motivated Strategies for Learning Questionnaire (MSLQ) of self regulated learning developed by Pintrich et al., (1991), revised by Johnson (2018) as MSLQ-R, and validated in the Indian context by Chechi, Bhalla and Chakraborty (2019), to extract the number of distinct profiles of the participant individuals. These sub-scales were critical thinking, organization and time and study environment of learning strategies scale, and self efficacy and intrinsic goal orientation of the motivation scale. The sample of the study comprised of 1799 undergraduate and post graduate university students of the state of Punjab, India from its three regions, Majha, Doaba and Malwa. The packages tidyLPA (2018, 2019) and dplyr of R/RStudio (2016) were used to conduct the study. The functions used were estimates_profiles, compare_solutions, plot_profiles, get_estimates and get_data. The models used in the study were the most lenient and the most strictest, model 1 and model 6, with variance and covariance, equal - zero and varying-varying in nature. The estimands used to resolve the number of profiles to extract were AIC, BIC, Entropy, BLRT-p-value. Three distinct profiles were extracted and they were identified as high (51 percent), average (35 percent) and low groups (14 percent) of individuals with self regulated learning as per the expectations. The R-codes to help in the replication of the study are provided. The significance of this research is discussed.

Keywords: *Latent Profile Analysis (LPA), Motivated Strategies for Learning Questionnaire (MSLQ), Indian students, tidyLPA, dplyr.*

INTRODUCTION:

One of the assumptions made by the researchers during the validation of any psychological instrument is that, all the participants have the same level of the measured construct in them. However, in reality, the participants differ from each other with respect to the presence of the measured construct in them. They can be broadly classified as belonging to certain number of homogeneous groups called profiles, where all the members of a profile are similar to each other with respect to the estimates of a certain estimands and differ with participants of other profiles. Such an outcome is statistically possible to be achieved through the general mixture model (Harring and Hodis, 2016; Pastor, Barron, Miller, and Davis, 2007) statistical technique called the Latent Profile Analysis (LPA), whose origin lies in developmental approaches (Magnusson and Cairns, 1996; Bergman and El-Khouri, 2003). The latent profile analysis is similar to factor analysis technique where a latent variable is assumed to exist which influences the variances in a host of manifest variables, and the technique aims to reduce the number of manifest variables into a manageable number of factors or dimensions. When the test items are replaced with the participants or persons and the latent dimensions are replaced with latent profiles, the approach becomes latent profile analysis, where it is assumed that beneath the obtained data from a sample, lies a set of homogeneous groups or profiles, into which each of the participant in the study can be associated with.

The variables used in the measurement of the parameters, like variances and covariances, using which the participants are classified, are continuous in latent profile analysis. If they are categorical, then

ISSN: 2233-7857 IJFGCN

another technique called the latent class analysis is conducted. Here, data of each of the continuous variables forms a distribution of its own, where every observation is assigned a probability for belonging to a sample or profile from a population of same dataset. From this heterogeneous general mixture model, a homogeneous group of participants is tried to be extracted by estimating the fit of data with certain pre-determined models, because of which latent profile analysis is called a model-based clustering technique (Hennig, Meila, Murtagh, and Rocci, 2015; Scrucca, Fop, Murphy, and Raftery, 2017).

There are six models of LPA which differ from each other based on the manner in which they measure or do not measure the parameters variance and co-variance, along with mean which is measured always. These six models are:

S.No.	Variance	Covariance	Model
1.	Equal	0	1 (Simplest)
2.	Varying	0	2
3.	Equal	Equal	3
4.	Equal	Varying	4 (in Mplus)
5.	Varying	Equal	5 (in Mplus)
6.	Varying	Varying	6 (Most Complex)

Models 4 and 5 can be estimated using MPlus software only. The rest of the four models can be estimated using the free-software package tidyLPA (2019) of R/RStudio (2016). The models 1 to 6 are arranged based on the increasing order of their complexity. The model 1 is the simplest and the most popular one. Here, only the mean of the participants from each profile is estimated. The variances of the each profile group are assumed to be same and the covariance is fixed at zero. The most complex model 6 takes into account the reality and hence makes no assumption. But, when the model is simple, it might not fit the data well and lack internal validity, but its external validity through the replication of results when applied using a fresh data set is high. When the model is complex, its internal validity is high as the data fits the model very well, but the changes of the replication of the estimates using afresh data comes down bringing down along with it the external validity of the data (Rosenberg et al., 2019). The steps involved in latent profile analysis can be listed as model specification, estimation of profiles, plotting the profiles, comparing of the solutions, getting the estimates of parameters and fitness of models.

Several types of fit indices estimates are used to find the model specification and its associated latent class, to finally estimate the number of profiles (Araujo et al., 2019). Some of the most commonly reported estimates are:

S.No.	Estimand	Meaning
1.	AIC: Aikake information criterion	Goodness of fit estimate which penalizes the model when its number of parameters increase. Lower the value, better the model fitness.
2.	BIC: Bayesian information criterion	Goodness of fit estimate which penalizes the model when its number of parameters increase, better than AIC. Lower the value, better the model fitness.
3.	SABIC: Sample size-adjusted Bayesian information criterion	Goodness of fit estimate which penalizes the model when its number of parameters increase, taking into consideration the sample size. Lower the value, better the model fitness.
4.	Entropy:	A measure of uncertainty in the classification of profiles, reverse-coded hence 1 means complete certainty in profile classification, and 0 means complete uncertainty
5.	Prob. Min.:	Lowest value of the diagonal of the average latent class probabilities for most likely class membership, as per the assigned profiles. It should be as high as possible, meaning that the cases are assigned to profiles which they must belong with a high probability
6.	Prob. Max.:	Greatest value of the diagonal of the average latent class probabilities for most likely class membership, as per the assigned profiles. It should be as high as possible, meaning that the cases are assigned to profiles which they must belong with a high probability

ISSN: 2233-7857 IJFGCN

7.	N Min.:	Depending on the most probable profile membership, the number of sample subjects assigned to the smallest profile
8.	N Max.:	Depending on the most probable profile membership, the number of sample subjects assigned to the largest profile
10.	BLRT p-value:	p-value for the bootstrapped likelihood ratio test. Significant p-value less than 0.05 represents goodness of fit between the model and the data.

The present study applied the latent profile analysis technique on the five sub-scales of MSLQ-R Indian version validated by (Chechi, Bhalla and Chakraborty, 2019), namely, organization, critical thinking, time and study environment from learning strategies scale and self efficacy and goal orientation from the motivation scale.

Table 1: Details of the Used MSLQ-R Sub-scales:

S.No.	Scale	Variable	Items
1.	"Motivation"	"Intrinsic Goal Orientation"	"1,16,22,24 (4)"
2.		"Self-Efficacy for Learning and Performance"	"5,6,12,15,20,21,29,31(8)"
3.	"Learning Strategy"	"Organization"	"32,42,49,63 (4)"
4.		"Critical Thinking"	"47,51,66,71 (4)"
5.		"Time Management and Study Environment"	"35,43, 65,70 (4)"

METHODOLOGY

Sample:

The sample of the study comprised of 1799 undergraduate and post graduates from the Majha, Malwa and Doaba regions of the Indian state of Punjab, belonging to the disciplines of Commerce, Science, Business Administration and Computer Application. Permission was taken from the head of the institutions to conduct the data questionnaire administration on the subjects during regular class sessions. The students were selected using simple random selection technique and they took 15-20 minutes to fill and return the questionnaire back to the investigator.

Statistical Analysis

The extraction of profiles as part of latent profile analysis is conducted using the *tidyLPA* package of R Ver. 3.6.3. along with the package *dplyr*. Model 1 where the variance is equal and covariance is zero and model 6 where both of them are varying were selected to estimate the profiles. Estimate_profiles and compare_solutions functions help in finding the optimum number of profiles. Help of estimands AIC. BIC, entropy and BLRT- p value were taken to finally settle for the number of profiles. The graphical representation of the profiles was presented by plot-profile function. The estimates of the profiles are obtained using get_estimates and get_data.

RESULTS

R Codes and Results of Latent Profile Analysis:

1. Import data file in r
2. > install.packages("tidyLPA")
3. > library(tidyLPA)
4. Install package dplyr
5. > library (dplyr)

ISSN: 2233-7857 IJFGCN

```
> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(1)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     1       27322.10 27377.05 1.00    1.00     1.00     1.00  1.00

> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(2)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     2       24460.22 24548.14 0.81    0.94     0.95     0.47  0.53  0.01


> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(3)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     3       23410.45 23531.34 0.82    0.88     0.93     0.14  0.52  0.01
> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(4)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     4       23028.69 23182.55 0.79    0.86     0.90     0.07  0.40  0.01
> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(5)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     5       22885.21 23072.04 0.80    0.68     0.91     0.04  0.39  0.01
> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(6)
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 1     6       22752.74 22972.54 0.76    0.71     0.91     0.05  0.32  0.01
```

Summary of Model 1 Specifications:

Model	Classes	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT_p
1	1	27322.1	27377.05	1.00	1.00	1.00	1.00	1.00	-
	2	24460.22	24548.14	0.81	0.94	0.95	0.47	0.53	**0.01**
	3	**23410.45**	**23532.34**	**0.82**	**0.88**	**0.93**	**0.14**	**0.52**	
	4	23028.69	23182.55	0.79	0.86	0.9	0.07	0.4	
	5	22885.21	23072.04	0.8	0.68	0.9	0.04	0.32	
	6	22752.74	22972.54	0.76	0.71	0.91	0.05	0.32	

ISSN: 2233-7857 IJFGCN

Interpretation: Though AIC and BIC estimates-wise, profile 3 is not the lowest, its entropy is highest at 0.82, which means that 82 percent of the cases of total 1799, that is 1475 cases, were properly classified into their most probable profile. 88 percent of the cases belonging to the lowest profile could be properly classified under this category as the Prob_min is 0.88. Since Prob_max is 0.93, it means that 93 percent cases belonging from the higher group were properly classified into its respective category. The number of cases in the lowest profile is 252 as the n-min is 0.14. The number of cases in the highest profile is 935. The rest of the cases comprising 34 percent, that is, 611 cases form the average group. The goodness of fit between the model and the data is very significant with p-value less than 0.05 at 0.01 of the estimand BLRT_p-value.

```
> MLSQ_SRL_Variables_Data %>% select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(3, variances = "varying", covariances = "varying")
tidyLPA analysis using mclust:

 Model Classes AIC      BIC      Entropy prob_min prob_max n_min n_max BLRT_p
 6     3       21999.31 22340.00 0.58    0.69     0.89     0.09  0.63  0.01
```

Summary of Model 1 and Model 6 Specifications:

Model	Classes	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT_p
1	3	23410.45	23532.34	0.82	0.88	0.93	0.14	0.52	0.01
6		21999.31	22340	0.58	0.69	0.89	0.09	0.63	0.01

Interpretation: When the number of parameters is high, the model estimation is the best. This is apparent since the AIC and BIC values of the model 6 estimating 3 profiles are less than the AIC and BIC values of the model 1 estimating 3 profiles. The entropy of the model 6 is very low though when compared to model 1. Both the model results are significant at 0.01 p-value of BLRT. A comparison of the estimates of classes 1,2 and 3 under model 1 and model 3 is shown below:

```
> MLSQ_SRL_Variables_Data %>% select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(1:3, variances = c("equal", "varying"), covariances = c("zero",
"varying"))%>%compare_solutions(statistics = c("AIC", "BIC"))
Compare tidyLPA solutions:

 Model Classes AIC      BIC
 1     1       27322.104 27377.053
 1     2       24460.225 24548.145
 1     3       23410.446 23531.335
 6     1       22753.962 22863.862
 6     2       22075.499 22300.794
 6     3       21999.310 22339.999

Best model according to AIC is Model 6 with 3 classes.
Best model according to BIC is Model 6 with 2 classes.
```


ISSN: 2233-7857 IJFGCN

An analytic hierarchy process, based on the fit indices AIC, AWE, BIC, CLC, and KIC (Akogul & Erisoglu, 2017), suggests the best solution is Model 6 with 3 classes.

```
> MLSQ_SRL_Variables_Data %>% select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(1:3, variances = c("equal", "varying"), covariances = c("zero",
"varying"))%>%compare_solutions(statistics = c("Entropy", "BIC"))
```

Compare tidyLPA solutions:

Model	Classes	Entropy	BIC
1	1	1.000	27377.053
1	2	0.806	24548.145
1	3	0.823	23531.335
6	1	1.000	22863.862
6	2	0.468	22300.794
6	3	0.582	22339.999

Best model according to Entropy is Model NA with NA classes.
Best model according to BIC is Model 6 with 2 classes.

An analytic hierarchy process, based on the fit indices AIC, AWE, BIC, CLC, and KIC (Akogul & Erisoglu, 2017), suggests the best solution is Model 6 with 3 classes.

Interpretation: All the estimates of the least strict (model 1) and the most strict (model 6) model specifications, show that the number of estimated classes or profiles for the present data is 3. They are termed as high SRL group, average SRL group and the low SRL group. The most popular model 1 estimates will be used for reporting the final results.

```
> MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(3)%>%plot_profiles()
```


ISSN: 2233-7857 IJFGCN

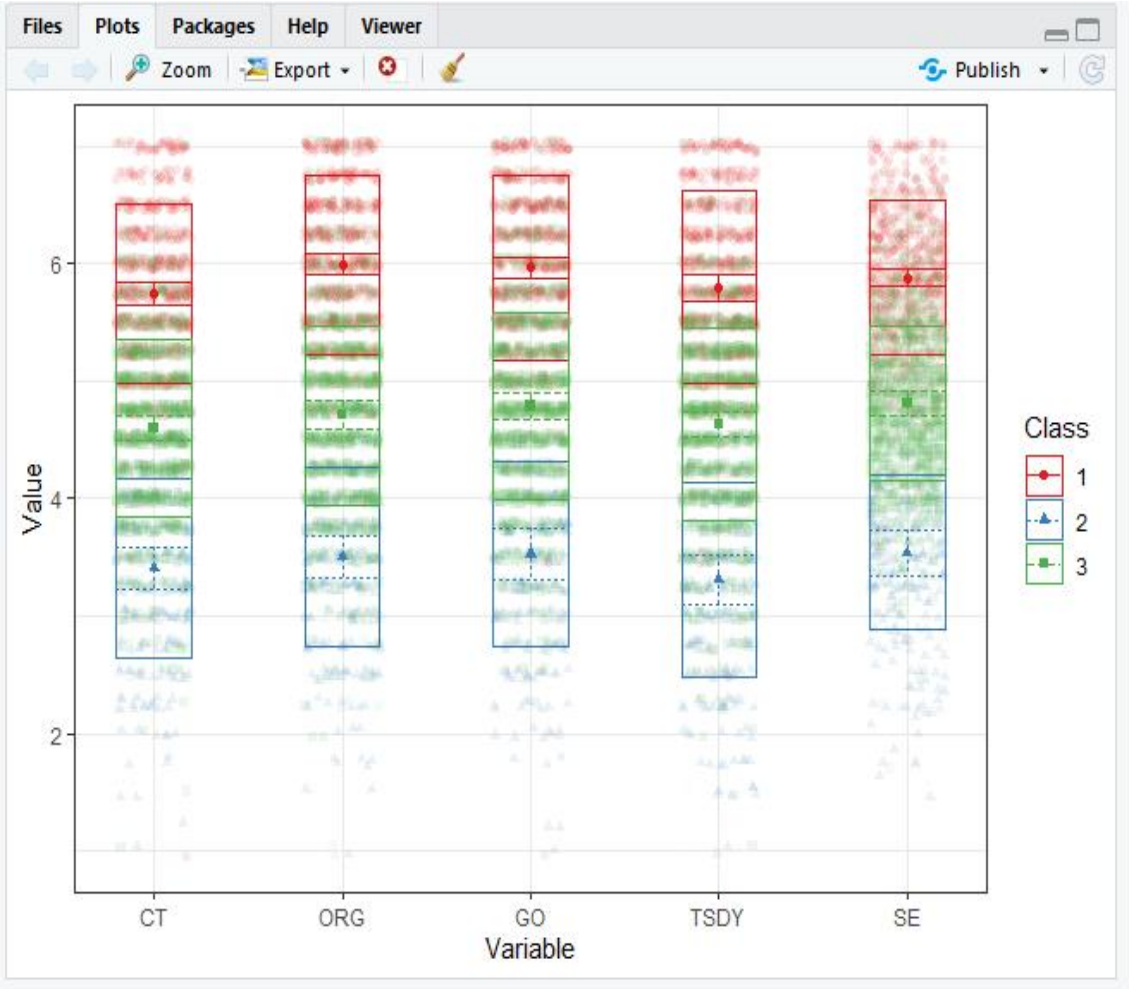

Fig.1: Latent Profiles of MSLQ-R

Interpretation: The low SRL group 2 in the above plot of profiles, is consistently low across all the variables of motivation and learning strategies sub-scales. The high SRL group 1, is consistently high across all the variables. Similarly, the average group represented by the profile 3, is so across all the variables of SRL.

```
> m <- MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>%
estimate_profiles(3)

>get_estimates(m)

# A tibble: 30 x 8
  Category  Parameter Estimate     se         p Class Model Classes
  <chr>     <chr>        <dbl>  <dbl>     <dbl> <int> <dbl>   <dbl>
 1 Means     CT         5.75  0.0501 0.          1     1       3
 2 Means     ORG          5.99  0.0441 0.          1     1       3
 3 Means     GO         5.97  0.0402 0.          1     1       3
 4 Means     TSDY         5.79  0.0507 0.          1     1       3
 5 Means     SE         5.88  0.0348 0.          1     1       3
 6 Variances CT         0.576 0.0278 3.61e- 95     1     1       3
 7 Variances ORG          0.590 0.0281 7.05e- 98     1     1       3
 8 Variances GO         0.626 0.0299 1.82e- 97     1     1       3
 9 Variances TSDY         0.675 0.0355 1.47e- 80     1     1       3
```


ISSN: 2233-7857 IJFGCN

```
10 Variances SE          0.429 0.0180 3.70e-125     1     1       3
# ... with 20 more rows
```

```
> m <- MLSQ_SRL_Variables_Data%>%select(CT,ORG,GO,TSDY,SE)%>%single_imputation() %>% estimate_pro
> get_estimates(m)
# A tibble: 60 x 8
   Category  Parameter Estimate        se     p Class Model Classes
   <chr>     <chr>        <dbl>     <dbl> <dbl> <int> <dbl>   <int>
 1 Means     CT          4.82   0.000658     0     1     1       1
 2 Means     ORG          4.97   0.000718     0     1     1       1
 3 Means     GO          5.01   0.000712     0     1     1       1
 4 Means     TSDY         4.84   0.000749     0     1     1       1
 5 Means     SE          4.99   0.000566     0     1     1       1
 6 Variances CT          1.18 NaN         NaN     1     1       1
 7 Variances ORG          1.29 NaN         NaN     1     1       1
 8 Variances GO          1.28 NaN         NaN     1     1       1
 9 Variances TSDY         1.35 NaN         NaN     1     1       1
10 Variances SE          1.02 NaN         NaN     1     1       1
# ... with 50 more rows
```

Interpretation: The estimates of the estimands for a particular profile under a specific model, or for a range of profiles can be obtained as shown above using the get_estimate function and a declared dataframe in r.

```
> get_fit(m)
# A tibble: 3 x 18
  Model Classes  LogLik    AIC    AWE    BIC   CAIC    CLC    KIC SABIC    ICL
  <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
1     1       1 -13651. 27322. 27480. 27377. 27387. 27304. 27335. 27345. -27377.
2     1       2 -12214. 24460. 24714. 24548. 24564. 24430. 24479. 24497. -24785.
3     1       3 -11683. 23410. 23761. 23531. 23553. 23368. 23435. 23461. -23887.
# ... with 7 more variables: Entropy <dbl>, prob_min <dbl>, prob_max <dbl>,
#   n_min <dbl>, n_max <dbl>, BLRT_val <dbl>, BLRT_p <dbl>
```

Interpretation: Similarly, the estimates of the goodness of fit for a particular profile under a specific model, or for a range of profiles can be obtained as shown above using the get_fit function and a declared dataframe in r.

Discussion:

Earlier, though the mathematical framework was in place for a number of statistical techniques, it was very cumbersome to manually compute them. With the availability of free open source softwares like R/RStudio (2016) and its large number of freely available packages to conduct variety of advanced statistical technique, there are hardly any pre-text left for the research community for not reporting the results as close to reality as possible now. In this context, it is the classification of subjects based on the extent of the presence of this measured variable in them was not much taken up, at least in the Indian context. It was believed that the construct was equally present in all the subjects. However, individual difference in psychological traits among the subjects is a reality. This short-coming of research is addressed well by the Latent profile analysis technique, and now made even more easier to conducted in

ISSN: 2233-7857 IJFGCN

computers using the tidyLPA and dplyr packages of R/RStudio off late. Until now the latent profile analysis was conducted only using the commercial software MPlus.

The availability of tidyLPA package should now lead to the proliferation of research studies using latent profile analysis of various constructs prost validation studies in the Indian context, to estimate the individual differences of the measured construct in the sample subjects.

Conclusion:

MSLQ has been used in multitude of SRL studies at post-graduate and doctoral level. However, the studies on profiling of the sample subjects remain scarce in the Indian context. The present is expected to serve as a hands-on tutorial for conducting this technique in the India.

REFERENCES

1. Araujo, A.M., Gomes, C.M.A., Almedia, L.S., & Nunez, J.C. (2019). A latent profile analysis of first-year university students' academic expectations, *Annals of Psychology*, 35(1), pp:58-67, doi: http://dx.doi.org/10.6018/analesps.35.1.299351
2. Bergman, L. R., & El-Khouri, B. M. (2003). A Person-Oriented Approach: Methods for Today and Methods for Tomorrow. *New Directions for Child and Adolescent Development*, 2003(101), 25–38.
3. Chechi, V.K., Bhalla, J. & Chakraborty, R. (2019). Cross Cultural Validation and Adaptation of the Parsimonious Version of Motivated Learning Strategies Questionnaire in the Indian Context, *International Journal of Advanced Science and Technology*, 28(16), ISSN: 2005-4238 , pp. 50-90.
4. Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist*, 51(3-4), 354-367.
5. Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of cluster analysis*. CRC Press.
6. Jackson, C., (2018). "Validating and Adapting the Motivated Strategies for Learning Questionnaire (MSLQ) for STEM Courses at an HBCU", *AERA Open*, Vol.. 4 no. 4, pp:1-16, DOI: 10.1177/2332858418809346.
7. Magnusson, D., & Cairns, R. B. (1996). *Developmental science: Toward a unified framework*. Cambridge, England: Cambridge University Press.
8. Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J.(1991). "A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
9. Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8-47. (https://www.sciencedirect.com/science/article/pii/S0361476X06000543
10. R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.
11. RStudio Team, RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, URL http://www.rstudio.com/, 2016.
12. RStudio Team, RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, URL http://www.rstudio.com/, 2016.
13. Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2018). tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. Journal of Open Source Software, 3(30), 978, https://doi.org/10.21105/joss.00978

ISSN: 2233-7857 IJFGCN

14. Rosenberg, J. M., van Lissa, C. J., Beymer, P. N., Anderson, D. J., Schell, M. J. & Schmidt, J. A. (2019). tidyLPA: Easily carry out Latent Profile Analysis (LPA) using open-source or commercial software [R package]. https://data-edu.github.io/tidyLPA/
15. Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2017) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, pp. 205-233

ISSN: 2233-7857 IJFGCN