

The Role of Personal Entry Characteristics in Dropout from Teacher Education

Doctoral Colloquium, 21 February 2024

David Simon

14 February 2024

Table of contents

Brief Introduction and Research Questions	1
Data Analysis	2
Further Considerations	8
Questions	9
References	9

Brief Introduction and Research Questions

Despite growing interest in dropout research, limited attention has been given to the context of teacher education. We utilized a Latent-Class-Analysis to investigate dropout, as this allows us to map sub-groups of students based on their individual characteristics prior to entering teacher education. In doing so we are able to take multiple characteristics and their interactions into account at the same time.

Evidence from general dropout research regularly highlights the relevance of individual characteristics prior to entering higher education, such as socioeconomic, familiar and educational background, personality or motivation for choosing a study subject. We are interested in the question which combination of prerequisites makes people disadvantaged by the conditions at higher education institutions. Our assumption is, that some groups of students do not succeed in integrating into the social or academic domain of higher education institutions and that, as a result, they will not be able to make productive use of the learning opportunities available and therefore be more likely to drop out of teacher education.

This concept is derived from the Student Integration Model (Tinto, 1975), one of the more influential theories aiming to explain dropout. It distinguishes between the social and academic integration of students and assumes that those who are not sufficiently integrated into the corresponding social and academic domain within higher education institutions decide to drop out. Academic integration encompasses academic performance as proof of formal affiliation as well as an adjustment to the norms held by higher education institutions themselves (e.g. achievement orientation). Social integration refers to the establishment of contacts with faculty and students, providing support and cognitive stimulation.

Both types of integration are seen to foster goal and institutional commitment, which we think will be reflected in an increased engagement and an increased utilization of learning opportunities available.

Consequently, we aim to answer three research questions:

Research Question 1: How many statistically distinct subgroups, regarding the entry characteristics of first-year teacher students, are there and can at least one of them be theoretically described as a risk group equaling an accumulation of unfavorable prerequisites in regards to educational failure?

Research Question 2: Do members of this group (these groups) show higher dropout rates compared to their peers?

Research Question 3: Can this relationship between risk group(s) and dropout from teacher education be explained through a failure of integration into the social or academic domain of higher education institutions and the subsequent unproductive utilization of learning opportunities?

This document is concerned with the first step of data analysis, the Latent-Class-Analysis, in order to address Research Question 1. If you are interested in the whole proposal you can have a look at the preregistration here: <https://doi.org/10.17605/OSF.IO/D2K7Y>.

I would like to ask you to read the following document. We had to make a number of decisions and I am interested in whether you find the approach convincing or not. The topics of interest are listed at the end of the document (section Questions).

Data Analysis

Variable Selection

Several determinants that define the dropout process within a general population of students prior to entering a study program have been distinguished (Heublein, 2014). Disadvantages are to be expected in the case of lower socioeconomic status, migration background, non-Gymnasium entrance certificate, lower prior academic achievements, a completion of a vocational training prior to entering higher education, lower scores on the personality trait conscientiousness and a primarily extrinsic motivation for choosing a degree program (Isleib et al., 2019). The set of entry characteristics discussed in this strand of research served us as a set of indicators for the Latent-Class-Analysis:

Variable name	
Parental education	1 = no parent has attained tertiary education (first generation students) 2 = one parent has attained tertiary education 3 = both parents have attained tertiary education
Parental occupation	1 = working class 2 = intermediate class 3 = upper class (EGP class scheme, the higher level of occupational class among both parents was used)
Migration background	1 = no, if both parents were born in Germany 2 = yes, if at least one parent was born abroad

Variable name	
Type of entrance certificate	1 = non-Gymnasium 2 = Gymnasium
Prior academic achievement (grade point average)	1 = 3.0–4.0 (“sufficient”) 2 = 2.6–3.5 (“satisfactory”) 3 = 1.6–2.5 (“good”) 4 = 1.0–1.5 (“very good”)
Big Five (extraversion, agreeableness, conscientiousness, neuroticism and openness)	10-item short version of the Big Five Inventory (Rammstedt & John, 2007), Response scale: 1 = does not apply at all 2 = does rather not apply 3 = does partly apply 4 = does rather apply 5 = does completely apply
Motivation for choosing teacher education (intrinsic motivation and extrinsic motivation)	Motivation for choosing teacher education questionnaire (Pohlmann & Möller, 2010), Response scale: 1 = not apply at that time at all 2 = rather not apply at that time 3 = did rather apply at that time 4 = did completely apply at the time

Data Base

Our analyses are based on data from the German Panel of Teacher Education as part of the Starting Cohort First-Year Students (SC5) of the National Educational Panel Study (NEPS) (Blossfeld & Roßbach, 2019). The NEPS initially included about 18000 first-year students in the winter term 2010/2011 and an oversampling of teacher education students (about 5500 students). From 2014 onwards, multiple survey instruments, for example, aspects of professional competence, instructional practices or professional development, were specifically addressed to teachers. The data was collected in 19 waves until 2022 (Schaeper, 2023).

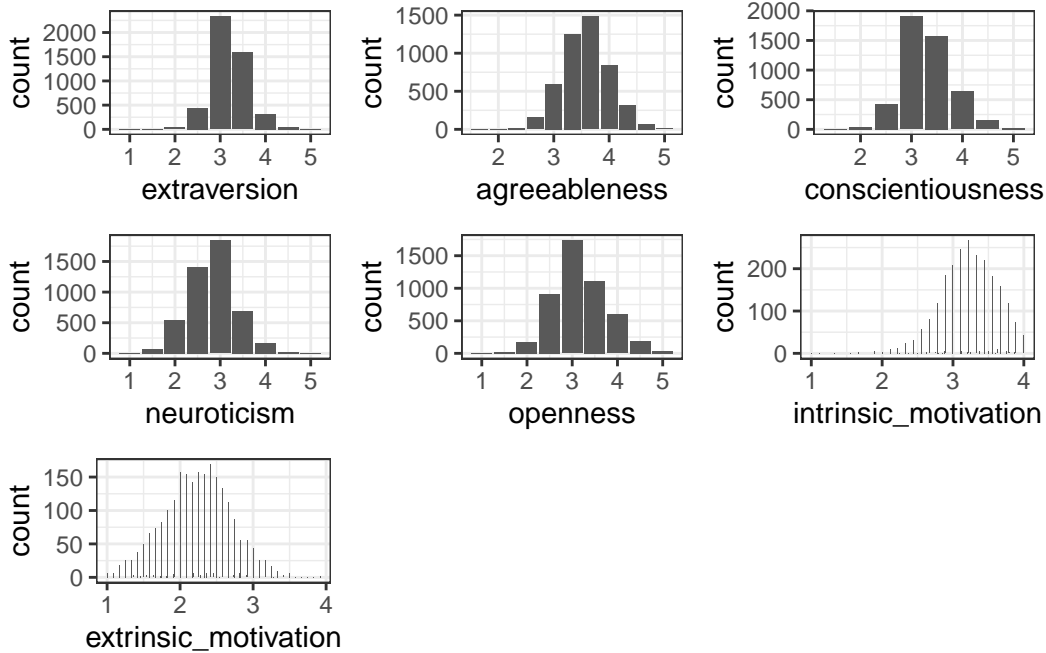
The resulting sample comprises **5775 first-year teacher students** ($M = 20.73$ years, $SD = 2.87$ years, 4337 women).

Examination of the Observed Data

While describing the data numerically, see table below, it is noticeable that the hypothetically continuous variables (Big 5 and motivation for choosing teacher education) do have just a few unique values and some response categories have near-zero frequencies. It is recommended that continuous variables should have “many” unique values and, if not, it is advised to better model them as ordinal (Van Lissa et al., 2023).

	name	type	n	missing	unique	mode
1	parental_education	numeric	5668	0.02	3	1.00
2	parental_occupation	numeric	4749	0.18	3	3.00

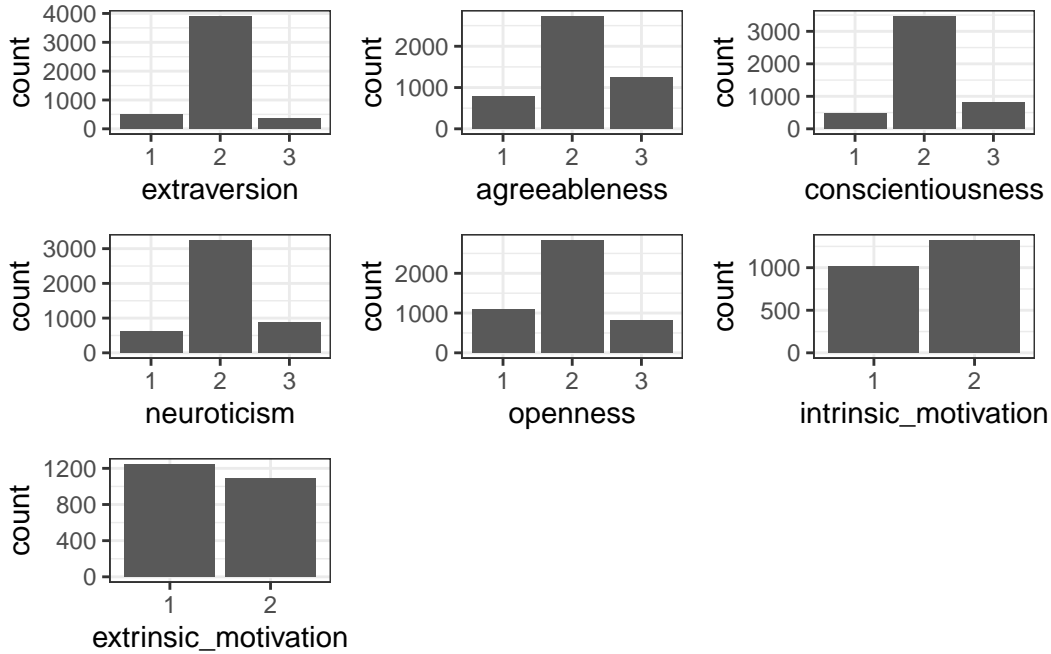
3	migration_background	numeric	5707	0.01	2	1.00
4	type_of_school	numeric	5757	0.00	2	2.00
5	prior_academic_achievement	numeric	5708	0.01	4	3.00
6	extraversion	numeric	4763	0.18	9	3.00
7	agreeableness	numeric	4763	0.18	11	3.67
8	conscientiousness	numeric	4763	0.18	8	3.50
9	neuroticism	numeric	4763	0.18	9	3.00
10	openness	numeric	4763	0.18	9	3.00
11	intrinsic_motivation	numeric	2341	0.59	44	3.22
12	extrinsic_motivation	numeric	2341	0.59	54	2.25



Consequently, we decided to rearrange these variables into categories. In order to retain the distributions of each variable we computed a three-level variable for the Big 5, including the categories low, medium and high. As a cut-off we used each variables tertile as there was a strong central tendency. In case of the motivation for choosing teacher education we took each variables median, dichotomizing the indicator into two categories: low and high. The procedure assumed that the presence of a risk factor is of interest in the present study, e.g. high levels of neuroticism. The aim was to be able to map such characteristics using this categorization.

	name	type	n	missing	unique
1	parental_education	factor	5668	0.02	4
2	parental_occupation	factor	4749	0.18	4
3	migration_background	factor	5707	0.01	3
4	type_of_school	factor	5757	0.00	3
5	prior_academic_achievement	factor	5708	0.01	5
6	extraversion	factor	4763	0.18	4

7	agreeableness factor	4763	0.18	4
8	conscientiousness factor	4763	0.18	4
9	neuroticism factor	4763	0.18	4
10	openness factor	4763	0.18	4
11	intrinsic_motivation factor	2341	0.59	3
12	extrinsic_motivation factor	2341	0.59	3



The variables shown above formed the input variables for further analyses.

Data Processing

We computed the Latent-Class-Analysis following the guide by (Weller et al., 2020) using the software poLCA (Linzer & Lewis, 2011). The package handles missing data by excluding any manifest variables with missing observations from the calculation. The priors within the expectation-maximization algorithm are updated in using as many or as few manifest variables as are observed for each individual (Linzer & Lewis, 2011). This follows the assumption that in computing LCA specific handling of missing values is not necessary.

Table 2: Results of the Latent -Class-Analysis (Model Fit Parameters)

model	log-likelihood	resid. df	BIC	aBIC	cAIC	likelihood-ratio	entropy
Modell 1	-41519	5754	83221	83154	83242	6636	
Modell 2	-41059	5732	82490	82353	82533	6264	0.48
Modell 3	-40950	5710	82464	82257	82257	6212	0.64
Modell 4	-40879	5688	82512	82235	82599	6165	0.43

model	log-likelihood	resid. df	BIC	aBIC	cAIC	likelihood-ratio	entropy
Modell 5	-40839	5666	82622	82276	82731	6145	0.38
Modell 6	-40800	5644	82735	82735	82866	6111	0.51

Note. resid.df = residual degrees of freedom, BIC = Bayesian Information Criterion, aBIC = adjusted BIC, cAIC = consistent Akaike Information Criterion.

Model fit was examined based on our theoretical understanding of dropout from teacher education and the BIC (Bayesian Information Criterion) as a statistical criterion, with lower BIC indicating better model fit (Nylund et al., 2007).

Additionally, we reviewed the following classification diagnostics criteria. We desired entropy above .80 and the lowest average latent class posterior probability to be between .80 and .90 (Weller et al., 2020).

We decided that the three-class solutions is to be preferred. It is the model with the lowest BIC and the highest entropy. Even if the latter and the lowest average latent class posterior probabilities are not as high as we desired, we interpret them as acceptable.

The average latent posterior probabilities are presented in a matrix with diagonals. It represents the average probability of a person being assigned to a class given the indicator variables used. Higher diagonal values are desirable. Off-diagonal elements in the posterior probability matrix contain probabilities of cases that belong in one class being assigned to another class (Bauer, 2021). The average latent class posterior probabilities of the three-class solution are displayed below:

Group.1	V1	V2	V3
1	1 0.71	0.04	0.26
2	2 0.05	0.89	0.06
3	3 0.15	0.01	0.83

Final Class Solution

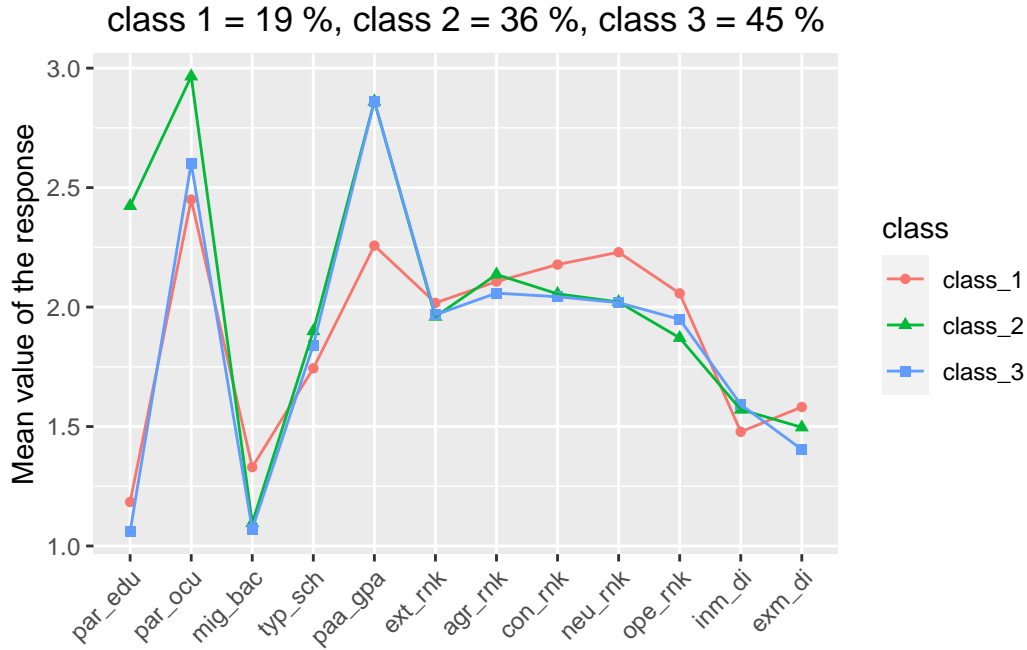


Figure 1. Latent Classes. par_edu = parental education (1, 2, 3), par_ocu = parental occupation (1, 2, 3), mig_bac = migration background (1, 2), typ_sch = type of school (1, 2), paa_gpa = prior academic achievement (1, 2, 3, 4), ext_rnk = extraversion (1, 2, 3), agr_rnk = agreeableness (1, 2, 3), con_rnk = conscientiousness (1, 2, 3), neu_rnk = neuroticism (1, 2, 3), ope_rnk = openness (1, 2, 3), inn_di = intrinsic motivation for choosing teacher education (1, 2), exm_di = extrinsic motivation (1, 2). Possible manifestations are displayed in brackets.

The classes can be described as follows:

Class 1 comprised 19 % of the sample.

Regarding social and familiar background variables this group of students is characterized through a high probability of being a first-generation student whose parents can be more likely assigned to working class. Students of this group had the highest probability of having a migration background and more likely attained their higher education entrance certificate not at a Gymnasium and with a low grade point average (most likely sufficient).

Regarding their personality traits, they were more likely highly extraverted, open and, contrary to what was assumed, showed the highest probability of being high in conscientiousness among all groups.

Regarding their motivation of choosing teacher education, members of this group described themselves most likely as high extrinsically motivated and low intrinsically motivated.

Class 2 describes 36 % of the sample.

In contrast to class 1, members of class 2 are described through the highest probability of both parents have attained tertiary education and are working under upper class regulations of employment. They

most likely have no migration background, attained a Gymnasium and finished school with most likely good grades.

Describing their personality it can be said that they most likely fall in the medium categories of extraversion, agreeableness, conscientiousness, neuroticism and openness.

They describe their motivation for choosing teacher education more likely to be intrinsically motivated and are ambiguous regarding extrinsic motives showing roughly equal conditional probabilities for both categories.

The **third** class compromised 45 % of the sample.

Regarding social and familiar background variables this group of students is, comparable to class 1, characterized through a marginalization due to a high probability of having a migration background and being a first-generation student. Their parents most likely can be assigned to intermediate class. But, in contrast to members of class 1, these students more likely attained their higher education entrance certificate at a Gymnasium and with a “good” grade point average - more comparable to class 2.

Also, regarding their personality traits, members showed similar distribution patterns as members of class 2 did.

Regarding their motivation of choosing teacher education, members of this group described themselves most likely as high intrinsically motivated and low extrinsically motivated.

Overall, class 2 shows conditional response probabilities regarding socioeconomic, familiar and educational background, personality and motivation that overall indicate positive effects in regards to academic success.

While class 3 seems to be at-risk in terms of their socioeconomic and familiar background they tend to have advantageous educational prerequisites as well as a favorable motivational state.

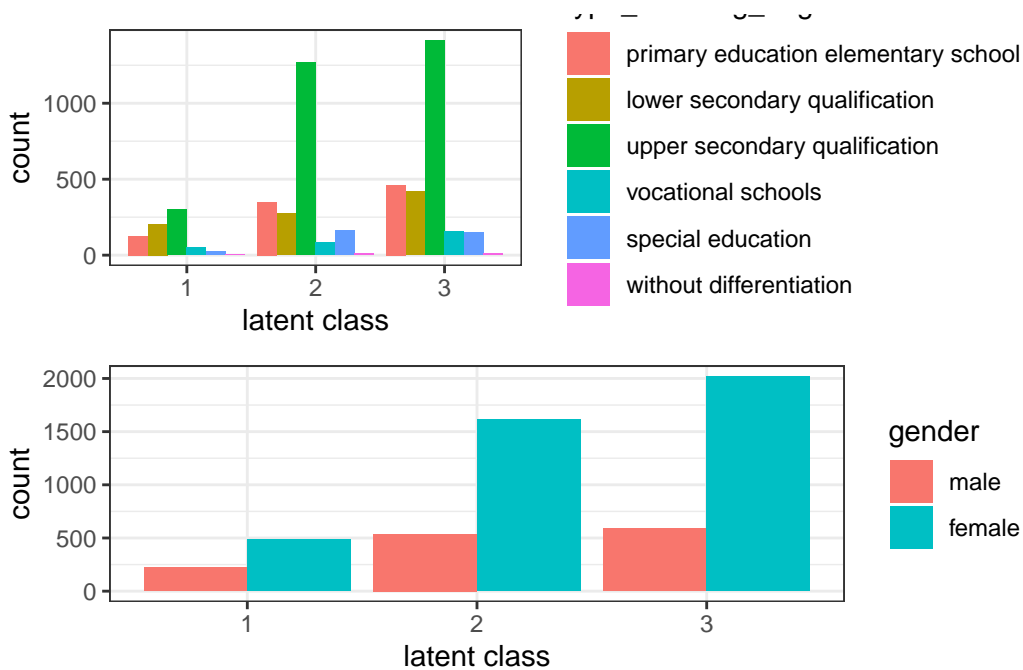
Class 1 showed itself potentially unfavorable on most variables and might be interpreted as a group of students potentially at-risk for educational failure and dropout. It must be noted that their high probability of being high in conscientiousness is contradicting.

Statistical tests on which conditional response probabilities differ from each other have not yet been computed. The portrayal is purely descriptive.

Further Considerations

In order to compute results on the other predictions, we would like to assign each case to a specific group based on their posterior class membership probabilities in order to use this as a manifest variable for further analyses. As this procedure is not uncontroversial and under suspicion of producing biased results (Bauer, 2021) we followed Sinha et al. (2020) and used a probability cut-off of 0.5 to assign class in order to address uncertainty.

To further describe the classes in a broader context, graphs showing the distribution of type of intended teaching degree as well as gender per class are displayed below.



Questions

- What do you think about the handling of the variables regarding personality and motivation for choosing teacher education?
- How would you rate the model selection?
- Is the graphical representation of the final model via the mean value of each response understandable, given the presence of formal categories?
- Do you know if Individual class assignment by modal posterior probability is (still) acceptable and a common practice?
- General considerations and thoughts are of course very welcome.

References

- Bauer, J. (2021). *A primer to latent profile and latent class analysis*. <http://dx.doi.org/10.31234/osf.io/97uab>
- Blossfeld, H.-P., & Roßbach, H.-G. (Eds.). (2019). *Education as a lifelong process*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-23162-0>
- Heublein, U. (2014). Student Drop-out from German Higher Education Institutions. *European Journal of Education*, 49(4), 497–513. <https://doi.org/10.1111/ejed.12097>
- Isleib, S., Woisch, A., & Heublein, U. (2019). Ursachen des Studienabbruchs: Theoretische Basis und empirische Faktoren. *Zeitschrift für Erziehungswissenschaft*, 22(5), 1047–1076. <https://doi.org/10.1007/s11618-019-00908-x>
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An r package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10), 1–29. <https://doi.org/10.18637/jss.v042.i10>

- 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. <https://doi.org/10.1080/10705510701575396>
- Pohlmann, B., & Möller, J. (2010). Fragebogen zur erfassung der motivation für die wahl des lehramtsstudiums (FEMOLA)1. *Zeitschrift Fur Padagogische Psychologie - Z PADAGOG PSYCHOL*, 24, 73–84. <https://doi.org/10.1024/1010-0652.a000005>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1), 203–212. <https://doi.org/https://doi.org/10.1016/j.jrp.2006.02.001>
- Schaeper, O., H. (2023). The german panel of teacher education students: Surveying (prospective) teachers from higher education into working life. *Journal of Open Psychology Data*, 11(1), 1–28. <https://doi.org/10.5334/jopd.76>
- Sinha, P., Calfee, C. S., & Delucchi, K. L. (2020). Practitioner’s Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. *Critical Care Medicine*, 49(1), e63–e79. <https://doi.org/10.1097/ccm.0000000000004710>
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89–125. <https://doi.org/10.3102/00346543045001089>
- Van Lissa, C. J., Garnier-Villarreal, M., & Anadria, D. (2023). Recommended Practices in Latent Class Analysis Using the Open-Source R-Package tidySEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–9. <https://doi.org/10.1080/10705511.2023.2250920>
- Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 46(4), 287–311. <https://doi.org/10.1177/0095798420930932>