Running Head : DEVELOPING A NEW LANGUAGE CLASS ANXIETY SCALE

Development and Validation of a New Multidimensional Language Class Anxiety Scale

Gulsah Kutuk[1], David W. Putwain[2], Linda Kaye[1], Bethan Garrett[1]

[1]Edge Hill University, Lancashire, UK

[2]Liverpool John Moores University, Liverpool, UK.

Address for correspondence: Gulsah Kutuk, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, United Kingdom. Email: Kutukg@edgehill.ac.uk

Abstract

This study reports on the development and assessment of a new 30-item Multidimensional Language Class Anxiety Scale which is designed to assess foreign language learners' anxiety regarding four language skills (listening, reading, writing and speaking) and testing. In Study 1, the initial items were piloted with 323 students studying English as a Foreign Language at three different universities in Turkey. This informed a revised version of the questionnaire which was subsequently administered to 701 students at three different Turkish universities. Confirmatory factor analyses revealed that a bifactor model with correlated residual variance yielded a better fit for the data in both studies than the other four models tested. The overall results provided preliminary evidence for the reliability and validity of the data collected using the new scale. Directions for future research and implications for foreign language teaching and learning are discussed.

Keywords: Foreign language anxiety, English as a foreign language; scale development; skill-based approach, psychoeducational assessment

Introduction

Within the field of second language acquisition (SLA), foreign language anxiety (hereafter, L2 anxiety) has been one of the primary concerns of researchers and practitioners for more than three decades (Dörnyei & Ryan, 2015). Considering previous research has also consistently reported that L2 anxiety is one of the most negatively influential psychological variables in L2 learning process (for a recent meta-analysis, see Teimouri, Goetze, & Plonsky, 2019), it is crucial for researchers to work toward developing a complete understanding of L2 anxiety adopting a variety of perspectives and approaches.

The idea of developing distinct situation-specific measures assessing L2 anxiety was first implemented by Horwitz et al. (1986). Horwitz et al. (1986) created the Foreign Language Classroom Anxiety Scale (FLCAS) which has generated a great deal of interest in language anxiety literature. Although it is still popular among the majority of the L2 anxiety researchers, some researchers raise the concern that the items in the FLCAS mainly focus on the speaking skill. This has been evidenced by several studies (Panayides & Walker, 2013; Park, 2014). Recent studies, however, have emphasized that it is not only speaking, but also the other three skills (i.e., listening, reading and writing) that need to be studied in L2 anxiety literature as well (Cheng, 2017). As Horwitz (2017) highlights, some learners might find listening, reading and writing more anxiety-provoking than speaking. As such, there is a need for studies focusing on the other language skills as well as speaking.

Furthermore, we need to acknowledge the importance of the role of test anxiety in language classrooms. It is one of the major components of L2 anxiety because almost all formal language learning and teaching contexts involve language testing and they play a significant role in learners' L2 achievement (Joy, 2013). Because test anxiety is not considered to be specific to L2 learning (Aida, 1994; MacIntyre & Gardner, 1989), there is

little research on L2 test anxiety and how it affects learners' L2 performance and there is no standardized measure specifically designed to assess L2 test anxiety (In'nami, 2006).

In light of the existing literature, the present study sought to develop a new multidimensional language class anxiety scale (MLCAS) which included the major domains of L2 anxiety (i.e., listening, speaking, reading, writing and testing) and differentiate between different components of anxiety (i.e., cognitive, affective and physiological). The need for a multidimensional skill-based L2 anxiety scale has recently been emphasized by Cheng (2017) whose measure also entails four brief scales measuring L2 learners' anxiety in relation to speaking, listening, reading, and writing. The current study is, therefore, timely and important to provide additional insights to the issue of measuring all the skills simultaneously. Differently from Cheng (2017), this study is specifically concerned with skill-based classroom activities and testing taking place in L2 classrooms.

Scale Development Procedure

To establish reliability and validity of data gathered using the MLCAS which is a 5-point Likert scale, two studies were conducted. The data gained from Study 1 and Study 2 were analysed using confirmatory factor analysis (CFA) with robust maximum-likelihood estimation (MLR) in M*plus* 7.4 (Muthen & Muthen, 2013). Traditionally, it is suggested that an exploratory factor analysis (EFA) needs to be done prior to a CFA to uncover the underlying structure of a set of latent constructs (Brown, 2014). However, we chose not to conduct an EFA on Study 1 data for two reasons. First, the factor structure was based on a strong theoretical basis, so it was known a priori. As suggested by Brown (2014), implementing a CFA is more appropriate than an EFA if researchers have a strong theory underlying the factor structure of their instruments. Second, an EFA would not be able to represent a potential complex bifactor structure of the MLCAS. In our case, in particular, similar item stem wording across the domains would create large method factors.

The MLCAS was developed using items adapted from the class-related anxiety scale of the Achievement Emotions Questionnaire (AEQ) (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Cronbach alpha was reported as $\alpha = .86$ for the class-related anxiety sub-scale of the AEQ (Pekrun, Götz, & Perry, 2005; Pekrun et al., 2011).

For each emotion including anxiety, Pekrun et al. (2005) generated items concerning affective, cognitive, physiological, and motivational components. The motivational component was excluded from the MLCAS to avoid the potential issues that might be caused by construct overlap with motivational antecedents and outcomes of language anxiety such as language motivation. The most suitable 2 items pertaining to cognitive, affective and physiological components were chosen on the basis of face validity. The procedure was repeated for all the skill-based anxieties and test anxiety and for three components of anxiety (Table 1). As the target samples consisted of Turkish students learning EFL, all the items were back-translated into Turkish.

<Insert Table 1 Here>

Study 1: Method

***Participants*:** Piloting was conducted among Turkish students studying EFL at university level. A total of 323 EFL students (male = 176, 45.5%; female = 147, 54.5%) with a mean age of 18.85 years (*SD* = 1.3) were recruited. There were no missing data regarding the variables needed for the study.

CFA. The validation of the data proceeded in two steps. First, based on the theory and logic behind language learning discussed in literature review, five plausible alternative models which could represent the structure of the MLCAS were proposed. Second, a CFA was conducted on the data gained from Study 1 to test the hypothesized factor structure of the models. The fit indices used to assess the models were the chi square (χ^2) statistic, degrees of freedom (df), the Root Mean Square Error of Approximation (RMSEA), the Standardized

Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), Tucker-Lewis index (TLI), Akaike Information Criterion (AIC) and sample-size adjusted Bayesian Information Criterion (aBIC). According to Hu and Bentler (1999), a good model is indicated by RMSEA < .05, SRMR < .08 and CFI and TLI > .95. As for AIC and aBIC, it is recommended that the model with the smallest value should be preferred (Hu & Bentler, 1999).

Results. The goodness of fit indices for all models are presented in Table 2. The CFA analysis revealed that Model 4, a bifactor solution with correlated residual variance, displayed adequate fit to the data when compared to the others. In this model, the correlated residuals were specified a priori. Without exception, all the residuals variances of the cognitive, affective and physiological items across different L2 domains were allowed to correlate (e.g., the residuals variances of the cognitive items in LAA, WAA, RAA, SAA and CTA). This was due to the fact that the items in these domains were closely related to each other. The results showed that the items in the MLCAS can be accounted by two processes: a single common factor that explains the common variance among 30 items and a set of factors that explain additional covariation among five sub-scales.

<Insert Table 2 Here>

Table 3 displays the standardized factor loadings, which provided preliminary support for the bifactor structure of the MLCAS.

<Insert Table 3 Here>

Descriptive Statistics. SPSS v.24 was used to generate descriptive statistics. Number of items in each construct, observed ranges, means, standard deviations, skewness and kurtosis. Internal consistency of the subscales as well as the overall scale was assessed using McDonald's omega (ω) which is suggested for its robust estimates (see Dunn, Baguley & Brunsden, 2014, McNeish, 2017). All target factors of the MLCAS yielded McDonald's ω scores ≥ .87 which meets the .70 cut-off criterion for reliability (Nunnally & Bernstein,

1994). The skewness and kurtosis statistics showed that all variables were normally distributed (Table 4).

<Insert Table 4 Here>

Study 2: Method

Participants. In Study 2, 3 different (1 private, 2 state) universities which were again based in Istanbul, Turkey were selected. Based on the convenience sampling method, 701 students (male = 346, 49.6%; female = 355, 50.6%) with a mean age of 19.17 years (*SD* = 1.9) were recruited.

CFA. To confirm the results obtained from Study 1, a second CFA was performed on the data gained from Study 2. This aimed to verify the factor structure, dimensionality, and internal consistency of the MLCAS.

Results. The goodness of fit indices for all models are presented in Table 5. The CFA results showed that Model 4, the bifactor model with correlated residual variance, provided a better fit to the main study data compared to the other models tested, which is consistent with the findings in Study 1. As highlighted in Study 1, correlated residual variances of cognitive, affective and physiological items across L2 domains were also included and allowed to correlate with each other in this model.

<Insert Table 5 Here>

Table 6 presents the standardized factor loadings for the subscales of the MLCAS. Factor loading estimates showed that the items substantially loaded onto their hypothesized factors (Table 6).

<Insert Table 6 Here>

Overall, the results from Study 2 confirmed that the MLCAS does not only consist of a single common factor that represents L2 class anxiety which is the multidimensional construct, but also addressed the five individual factors which are LAA, WAA, RAA, SAA

and CTA that compromise it. It also accounts for the variance due to the different components of anxiety which are cognitive, physiological and affective (Figure 1).

<Insert Figure 1 Here>

Assessment of Internal Consistency. Consistent with Study 1, all the sub-scales as well as the overall scale showed good internal consistency (McDonald's ω ranging from .88 to .96). All the factors were normally distributed (Table 7).

<Insert Table 7 Here>

Predictive validity of the MLCAS. We also explored whether there is a link between the participants' overall language performance scores and the MLCAS results. We ran a latent correlation analysis in M*plus* using the second study data to examine correlations between participants' overall language performance scores and the subscales of the MLCAS along with the overall scale. The performance scores were measured by the language tests taken by the participants after completing their one-year English language programme. The tests included all the four language skills namely reading, writing, listening and speaking. The results showed that all the variables were significantly and negatively correlated with the performance scores (see Table 8).

<Insert Table 8 Here>

Discussion

This study has three major contributions to L2 anxiety research. Using a concise scale measuring L2 anxiety pertaining to all the language skills and testing simultaneously will be of great support to L2 anxiety researchers as they can measure L2 anxiety without compromising the length of the scale. Secondly, unlike the items in many other scales in the L2 anxiety literature, the items which were adapted from the AEQ (Pekrun et al., 2005) offer a unique and contemporary approach to measuring emotions such as anxiety. The MLCAS differentiated between affective, cognitive and physiological components of anxiety

enhancing our understanding of emotions in two ways. Theoretically, it offers a better understanding of the nature of L2 anxiety which L2 learners suffer from. Practically, it makes it possible for language teachers to help anxious L2 learners appropriately. Lastly, as a further contribution, we developed a measure of L2 test anxiety which is intended to correspond as an additional sub-scale on the MLCAS. In this study, it has been shown that test anxiety is indeed a part of L2 anxiety and it should be assessed along with the language skills. Therefore, it could be possible to differentiate whether learners are in fact anxious because of a particular language skill (listening, reading, writing, and speaking) or it is just the nature of testing that makes them worried.

However, the findings presented here are provisional and should be treated cautiously until more research has been conducted to replicate the results. Although there were 1024 participants involved in this research, they were all Turkish learning English as a foreign language. Therefore, it is not possible to ascertain the extent to which the findings presented here can be generalized to other language learning contexts. More research with different groups of learners and languages other than English is needed to get further support for the generalizability of the results gained using the MLCAS. Also, it is important to note that the wording of the items used in the scale was similar to each other, which is another limitation of the current study. Therefore, it is suggested that any future work on the measure should reduce item similarity going forward.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal, 78*, 155-168. doi:10.1111/j.1540-4781.1994.tb02026.x

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* New York, NY: Guilford.

Cheng, Y. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System, 68,* 15-25. doi:10.1016/j.system.2017.06.009

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York, NY: Routledge. doi:10.4324/9781315779553

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399-412. doi:10.1111/bjop.12046

Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz, and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney & J. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31-47). Bristol, UK: Multilingual matters.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*, 125-132. doi:10.1111/j.1540-4781.1986.tb05256.x

Hu, L., & 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. doi:10.1080/10705519909540118

In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System, 34*, 317-340. doi:10.1016/j.system.2006.04.005

Joy, J. L. (2013). The altitude of test anxiety among second language learners. *Language Testing in Asia, 3*(1), 10. doi:10.1186/2229-0443-3-10

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning, 39*, 251-275. doi:10.1111/j.1467-1770.1989.tb00423.x

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*, 412. doi:10.1037/met0000144

Muthén, L. K., & Muthén, B. O. (2014). M*plus user's guide* (7th ed.). Los Angeles, CA: Author.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychological theory*. New York, NY: MacGraw-Hill.

Panayides, P., & Walker, M. J. (2013). Evaluating the psychometric properties of the foreign language classroom anxiety scale for cypriot senior high school EFL students: The rasch measurement approach. *Europe's Journal of Psychology, 9*, 493-516. doi:10.5964/ejop.v9i3.611

Park, G. (2014). Factor analysis of the foreign language classroom anxiety scale in korean learners of english as a foreign language. *Psychological Reports, 115*, 261-275. doi:10.2466/28.11.PR0.115c10z2

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36-48. doi:10.1016/j.cedpsych.2010.10.002

Pekrun, R., Götz, T., & Perry, R. P. (2005). Achievement emotions questionnaire (AEQ). user's manual. *Department of Psychology, University of Munich.*

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. Studies in Second Language Acquisition, 41, 363-387. doi:10.1017/S0272263118000311

Table 1

An Example of the Adapted Versions of an Affective Item from the Achievement Emotions Questionnaire

Original item	I feel nervous in class
LAA	I feel nervous during *listening* activities in English class
WAA	I feel nervous during *writing* activities in English class.
RAA	I feel nervous during *reading* activities in English class.
SAA	I feel nervous during *speaking* activities in English class.
CTA	I feel nervous while taking an English language *test*.

Note: Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Listening Activity Anxiety (LAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA)

Table 2

The goodness of fit indices for all models – Study 1

Model	Number of factors	χ^2	*df*	RMSEA	CFI	TLI	SRMR	AIC	aBIC
1	5 lower order factors	567.28***	260	.061	.944	.907	.099	22732.70	22870.67
2	5 lower order factors and 1 higher order factor	607.02***	265	.064	.938	.898	.103	22776.52	22911.56
3	5 lower order factors and 3 method factors	657.39***	362	.051	.946	.936	.047	22632.64	22710.73
4	Bifactor model with correlated residual variance	309.69***	230	.033	.986	.973	.027	22459.07	22614.66
5	Bifactor model without correlated residual variance	777.23***	365	.058	.929	.915	.051	22757.83	22834.16

***$p < .001$.

Table 3

Standardized Loadings for the five lower order factor solution with three method factors – Study 1

Items	Factors					
	RAA	WAA	LAA	SAA	CTA	LCA
1. R1	0.419					0.810
2. R2	0.557					0.775
3. R3	0.070					0.728
4. R4	0.194					0.753
5. R5	0.345					0.762
6. R6	0.002					0.748
7. W1		0.688				0.419
8. W2		0.568				0.684
9. W3		0.314				0.696
10. W4		0.595				0.396
11. W5		0.588				0.572
12. W6		0.079				0.666
13. L1			0.467			0.777
14. L2			0.458			0.740
15. L3			0.158			0.790
16. L4			0.251			0.542
17. L5			0.443			0.786
18. L6			0.144			0.534
19. S1				0.425		0.586
20. S2				0.465		0.613
21. S3				0.213		0.759
22. S4				0.480		0.619
23. S5				0.681		0.599

24. S6	0.133		0.716
25. T1		0.599	0.631
26. T2		0.692	0.626
27. T3		0.202	0.512
28. T4		0.475	0.555
29. T5		0.594	0.579
30. T6		0.353	0.586

Note: Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Listening Activity Anxiety (LAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA), L2 Class Anxiety (LCA)

Table 4

Item and Scale Statistics – Study 1

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	McDonald's omega (ω)
LAA	6	6-30	6-30	14.43	5.61	.297	-.657	.90
SAA	6	6-30	6-30	15.26	5.95	.207	-.673	.90
RAA	6	6-30	6-30	12.94	5.50	.622	-.073	.91
WAA	6	6-30	6-29	14.21	5.17	.352	-.487	.88
CTA	6	6-30	6-30	16.58	5.73	.009	-.640	.89
LCA	30	30-150	30-143	73.95	24.44	.156	-.513	.96

Note: Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Listening Activity Anxiety (LAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA), L2 Class Anxiety (LCA)

Table 5
The goodness of fit indices for all models – Study 2

Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR	AIC	aBIC
1	776.00***	261	.058	.946	.911	.077	41766.31	42046.40
2	798.737***	266	.058	.945	.909	.075	41787.76	42061.87
3	2137.71***	375	.090	.815	.786	.343	42978.68	43121.50
4	482.180***	241	.042	.975	.954	.030	41156.92	41459.21
5	1231.13***	375	.063	.910	.896	.059	41841.94	41984.75

***$p < .001$.

Table 6

Standardized factor loadings for the bifactor model with correlated residual variance – Study 2

Items	Factors					
	RAA	WAA	LAA	SAA	CTA	LCA
1. R1	0.562					0.621
2. R2	0.376					0.528
3. R3	0.284					0.718
4. R4	0.410					0.651
5. R5	0.484					0.733
6. R6	0.051					0.656
7. W1		0.595				0.588
8. W2		0.633				0.490
9. W3		0.301				0.550
10. W4		0.499				0.622
11. W5		0.459				0.715
12. W6		0.201				0.804
13. L1			0.640			0.523
14. L2			0.332			0.757
15. L3			0.563			0.588
16. L4			0.351			0.716
17. L5			0.169			0.795
18. L6			-0.054			0.854
19. S1				0.700		0.589
20. S2				0.688		0.604
21. S3				0.317		0.609
22. S4				0.321		0.514
23. S5				0.578		0.537
24. S6				0.020		0.828
25. T1					0.578	0.580
26. T2					0.527	0.699

27. T3	0.247	0.622
28. T4	0.463	0.540
29. T5	0.423	0.711
30. T6	0.239	0.741

Note: Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Listening Activity Anxiety (LAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA), L2 Class Anxiety (LCA)

Table 7
Item and Scale Statistics for Model 4 – Study 2

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	McDonald's omega (ω)
SAA	6	6-30	6-30	16.42	5.59	.109	-.613	.88
LAA	6	6-30	6-30	14.97	5.55	.261	-.266	.89
RAA	6	6-30	6-30	13.60	5.03	.559	-.301	.88
WAA	6	6-30	6-30	14.21	5.29	.495	-.007	.89
CTA	6	6-30	6-30	17.85	5.90	-.220	-.617	.89
LCA	30	30-150	30-147	77.05	23.58	.081	-.099	.96

Note: Reading Anxiety (RA), Writing Anxiety (WA), Listening Anxiety (LA), Speaking Anxiety (SA), Test Anxiety (TA), L2 Anxiety (LLA)

Table 8
Correlation analysis

		Performance
1.	SAA	-.189***
2.	RAA	-.298***
3.	WAA	-.266***
4.	LAA	-.202***
5.	CTA	-.243***
6.	LCA	-.269***

***$p < .001$.

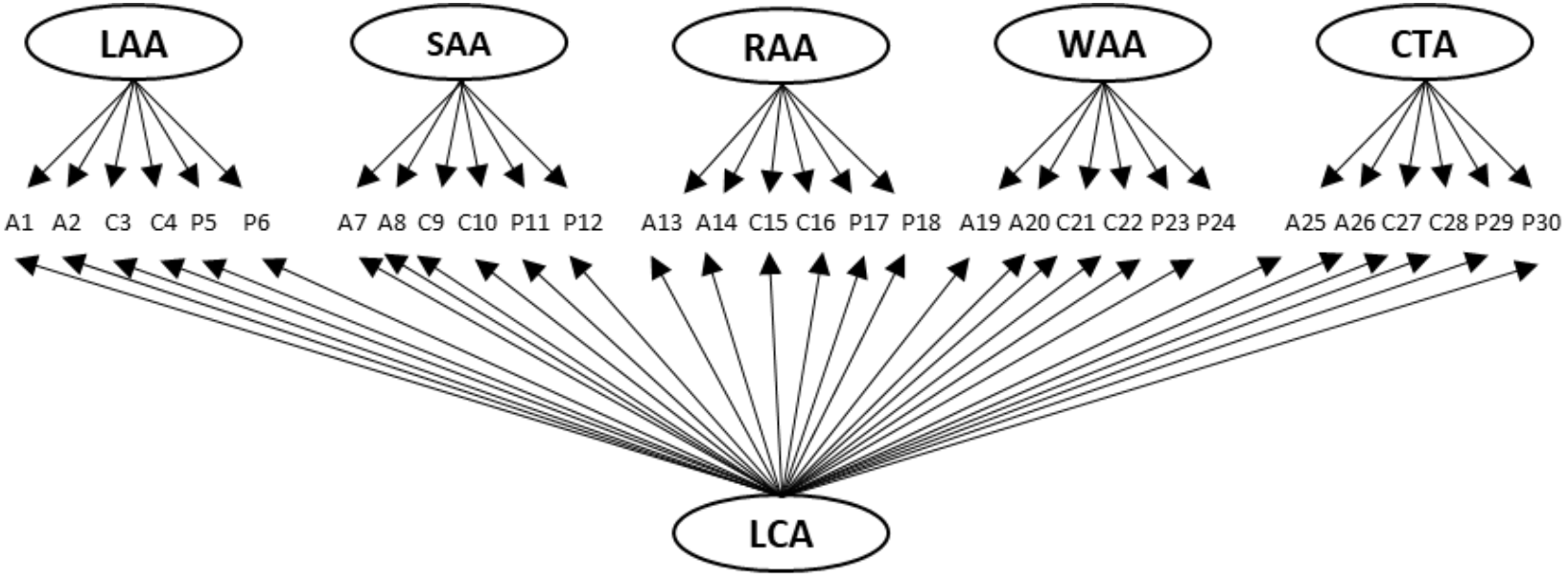

Figure 1. Bifactor model with correlated residual variance. For simplicity, the relations between corresponding residuals were omitted.