THE EUROPEAN EDUCATIONAL RESEARCHER
DOI: 10.31757/euer.435

http://www.eu-er.com/

Turkish Adaptation and Psychometric Evaluation of the Colorado Learning Attitudes About Science Survey (CLASS) in Physics

Derya Kaltakci-Gurel
Kocaeli University, Turkey

Abstract: The Colorado Learning Attitudes about Science Survey (CLASS) is an instrument to measure student beliefs about physics and learning physics. In this research, Turkish adaptation and psychometric evaluation of the CLASS is discussed. In the first stage, the translation process, which included examination of six experts (four experts in physics education and two experts in English and Turkish languages) for content validity and 13 student interviews for face validity, was described. In the second stage, exploratory (EFA) and confirmatory (CFA) factor analysis results obtained from 1391 freshman students were discussed for construct validty. The EFA yielded three factors that consisted of 20 items, which explained 39.61 % of the total variance. These factors were named as: *Problem Solving Effort, Conceptual Understanding, and Personal Interest and Real-World Connection.* Based on the CFA results, the three-factor 20-item instrument showed acceptable fit statistics. Compared to the original CLASS, the proposed version with 20-item model was shorter, easier to administer and easier to score, valid and reliable survey, so feasible to use. The overlapping of the constructs (i.e., violation of the unidimensionality) in the original CLASS was a problem, while in the proposed model none of the items were included in more than one construct.

Keywords: *Attitudes; Beliefs; CLASS; Physics Education.*

Introduction

Over the last decades, numerous assessment instruments have been developed to investigate and assess student attitudes and beliefs in physics. Some of these instruments are the Views About Science Survey (VASS) (Halloun, & Hestenes, 1998), the Views of Nature of Science (VNOS) (Lederman, Abd-El-Khalick, Bell, & Schwatz, 2002), the Maryland Physics Expectation Survey (MPEX) (Redish, Saul, & Steinberg, 1998), the Epistemological Beliefs Assessment for Physical Sciences (EBAPS) (Elby, Fredriksen, Scwarz, & White, 2006), and the Colorado Learning Attitudes about Science Survey (CLASS) (Adams, Perkins, Podolefsky, Dubson, Finkelstein, & Wieman, 2006).

The quality of the assessments used in research determines the advancement of knowledge in physics education (Douglas, Yale, Bennett, Haugan, & Bryan, 2014). Effective instrument development requires a systematic, well-organized approach to ensure sufficient validity and reliability evidence to support the proposed inferences from the scores (Downing, 2006). Validity is the adequacy and appropriateness of the interpretations and uses of assessment results. Since validity is a matter of degree, not all or none, there exists no absolute method to determine the validity of the test scores. However, qualitative and quantitative evidence from different sources should be collected to ensure validity. Reliability refers to the consistency of the scores and the extent to which the measures are free from errors. We can measure error by estimating how consistently a trait is assessed (McMillan & Schumacher, 2001; Linn, 2006). According to Classical Test Theory statistics obtained for validity and reliability evidence are always sample dependent. So an instrument validated for one population may not be valid for another, necessitating the need to continue validating an instrument. Replicating a study provides evidence of the trustworthiness of the original results (Douglas et al., 2014).

The CLASS is one of the most widely used surveys in physics education research to explore students' epistemological beliefs about physics and physics learning since its first development in 2006. The CLASS builds on the existing surveys (MPEX, VASS, EBAPS) and adds information to account for other student attitudes and beliefs observed to be important in educational practice (Adams et al., 2006). Since its development, the CLASS has been used by several researchers for different purposes (Alhadlaq, Alshaya, Abdulkareem, Perkins, Adams, & Wieman, 2009; Bates, Galloway, Loptson, & Slaughter, 2011; Bayar & Karamustafaoglu, 2015; Kost, Pollock, & Finkelstein, 2009; Lindsey, Hsu, Sadaghiani, Taylor, & Cummings, 2012; Milner-Bolotin, Antimirova, Noack, and Petrov, 2011; Perkins, Adams, Pollock, Finkelstein & Wieman, 2004; Zhang, & Ding, 2013): including making distinctions between beliefs of experts compared to beliefs of novices; investigating the changing attitudes during time (from the beginning to the end of a semester); researching the differences in attitudes and beliefs according to gender, country or teaching methods used; or the relationship between attitudes and achievement. However, the psychometric evaluations of the CLASS were done by only a few researchers (Bayar & Karamustafaoglu, 2015; Douglas et al., 2014; Heredia & Lewis, 2012; Kontro & Buschhüter, 2020). By means of giving the instrument to a new sample to confirm that the instrument behaves as predicted is valuable to continue validating an instrument. Similar psychometric analysis was suggested for different samples, cultures and versions of CLASS (Heredia & Lewis, 2012). In the present study, Turkish translation and validation of the CLASS for university students was discussed in detailed.

Although the CLASS has been one of the most widely used surveys in physics education research for several purposes, the survey validation studies are scarce. Sawtelle, Brewe and Kramer (2009) conducted a validation study of the CLASS on a minority population. They conducted student interviews with 30 Hispanic students in Florida International University as a survey validation method, and concluded that except one item (item 21) the CLASS survey is valid for use with Hispanic student population. In another study, Douglas et al. (2014) administered the CLASS to a sample of 3844 university students in the USA and investigated the psychometric properties of the survey. They performed several iterations and obtained some models based on their factor analysis results. Their final model with 17 items and 15 items were re-evaluated together with model fit indices. As a result, they proposed a 15-item instrument with three factors to best fit for their data taken from university students. More recently, Kontro and Buschhüter (2020) conducted another validation study on 642 Finnish university students, who were culturally and demographically different from the aforementioned two studies. They re-evaluated the 3-factor 15 items model proposed by Douglas et al. (2014), and reported a model similar to Douglas et al. (2014), but excluded item 25, providing the best description of their data. They also extended their statistical validity analysis through factor analysis to expert ratings and student interviews, and discovered that the results to support each other.

Instrument-The CLASS

The CLASS survey, which was developed, validated, and administered, by Adams, Perkins, Podolefsky, Dubson, Finkelstein and Wieman (2006) at the University of Colorado in the U.S., is a survey of Likert-type items. The survey mainly designed to measure student beliefs about physics and about learning physics was administered to a wide range of students and physics faculty. The CLASS was originally developed in English, and translated into several other languages including Arabic, Chinese, Finnish, German, Japanese, Portuguese, Spanish, Swedish, and Turkish[1]. Since its development, the CLASS survey has been used in numerous studies published in physics education research. Also, it has been modified for biology and chemistry, which are available online[2]. In the CLASS students were asked to respond on a 5-point likert-type (from strongly disagree to strongly agree) scale to statements such as: "Knowledge in physics consists of many disconnected topics". It has been used in both pencil and paper and online formats. The 42-item CLASS was first developed and tested for over 5000 students and physics faculty experts. Based on student responses eight categories were empirically determined by the researchers. These eight categories were: 1) *Real World Connection*, 2) *Personal Interest*, 3) *Sense Making/Effort*, 4) *Conceptual Connections*, 5) *Applied Conceptual Understanding*, 6) *Problem Solving General*, 7) *Problem Solving Confidence*, 8) *Problem Solving Sophistication*. However, from 42 items in the survey only 26 of them were assigned in these eight categories, and many of these items overlapping between categories. That means, those categories were not unique since an item could be included in more than one category. For example, as shown in Table 5 all of the items in three of the categories (*Conceptual Connections, Problem Solving Confidence, Problem Solving Sophistication*) were included in any of the other categories. Also, there were at least two items in each category that also included in any of the other categories. To determine students who randomly choose answers in the survey, item 31 in the survey was used as a control item.

Adams et al. (2006) performed a series of validation and reliability studies to revise and refine survey items. For face validity, they made interviews with physics faculty for expert opinion and with students to confirm clarity of items. For construct validity, they performed factor analysis to create and verify categories of items. A correlation with student beliefs and course performances were used for predictive validity. For concurrent validity, they showed that physics majors are more expertlike in their beliefs than nonscience majors. However, as discussed in several psychometric reevaluation studies (Douglas et al., 2014; Heredia & Lewis, 2012; Kontro & Buschhüter, 2020) of the CLASS, Adams and colleagues' proposed solution said to violate the unidimensionality of each construct (i.e., category) in terms of validity. In their psychometric evaluation study Heredia and Lewis (2012) proposed a 16-item, three-factor solution to best fit their data in CLASS-Chemistry version. Douglas et al. (2014) performed a similar psychometric evaluation for the CLASS-Physics version. They proposed a 15-item, three factor solution produced the acceptable fit statistics. These three factors were *Personal application of physics* (items 3,14,25,28,30,37), *Personal effort in a physics course* (items 23, 24, 29, 32), and *Approaches to problem solving* (5, 21, 22, 34, 40). Kontro and Buschhüter (2020) validated the three-factor model of Douglas et al. (2014) over their Finnish student data obtained at the University of Helsinki by confirmatory factor analysis and adjusted the model to fit indices. They proposed a 14-item (excluding item 25), three-factor solution produced the acceptable fit statistics. As Douglas et al. (2014) suggested, researchers must continuously re-examine the psychometric properties of popular assessments, such as CLASS, on different groups and modifications to the survey, if any, should be shared with other researchers, since validity is an ongoing process.

Therefore, the purposes of this study were: (1) to describe the process of Turkish adaptation of Colorado Learning Attitudes about Science Survey (CLASS) in Physics, and (2) to examine the psychometric properties of Colorado

Learning Attitudes about Science Survey (CLASS) in Physics, and propose a model based on statistical analysis to improve the validity of the interpretation and administration of the CLASS results.

Methods

Research Design

In this research a cross-sectional survey design was used. In this design the aim was to collect information about students' beliefs about physics and learning physics at a just one point in time.

Sample

In this study, the CLASS was translated and adapted into Turkish. For the translation process interviews were conducted with 13 freshman students. These 13 participant students were selected from physics education (PHED) and chemistry education (CHED) programs for the ease of access to the individuals. The maximum variation sampling method was used to select participants from these two departments. The maximum variation was done according to the Cumulative Grade Point Averages (CGPA) of students. High, medium and low achieving categories were determined according to CGPAs and an effort was made to select participants for the interviews equally from those three categories. As a result, the interviews were conducted with 5 (38,5 %) physics education, 8 (61,5 %) chemistry education program students; 8 (61,5 %) were female and 5 (38,5%) were male.

The CLASS in Turkish was administered as a paper-pencil format during 2010-2011 academic year in a university from the middle part of Turkey. The university was selected by convenience for the ease of administration of the survey. The university from which the sample selected is one of the top ranking state universities in Turkey. The university is a technical university such that 23 of the departments in the university have physics courses as compulsory in their curriculum in first year of a four-year program. A total of 1391 students from 22 departments were taken the survey during their second semester General Physics Laboratory II class in 10-15 minutes. Only one of the departments (Elementary Mathematics Education-EME) could not be included in the sample because of random problems. The researcher was available in the classrooms during the administrations of the survey together with the instructors to inform the participants about the survey and to direct and observe the process. During the preliminary analysis of the data obtained from 1391 subjects, 171 subjects (12 %) were removed from further analysis of the study. These 171 subjects were either select incorrect option in the control item (item 31) in the CLASS or had omitted response more than 10 % of the survey items. Therefore, all statistical analysis was performed over these 1220 subjects (38 % female, 62% male). The subjects' demographic details are presented in Table 1.

Table 1

Distribution of Sample Subjects to Their Departments and Gender.

	Gender		Total Number and Percentages of Subjects	
Department	**Female**	**Male**	**N**	**%**
AEE	15	36	51	4.2
BIO	24	10	34	2.8
CE	18	105	123	10.1
CEIT	23	27	50	4.1
CENG	18	68	86	7.0
CHE	54	24	78	6.4
CHED	13	5	18	1.5
CHEM	28	10	38	3.1
EE	23	117	140	11.5
ENVE	18	8	26	2.1
ESE	31	9	40	3.3
FDE*	29	10	40	3.3
GEN	15	6	21	1.7
GEO	12	26	38	3.1
IE	26	33	59	4.8
MATH	31	19	50	4.1
ME*	25	108	134	11
METE	23	28	51	4.2
MINE*	10	26	37	3.0
PETE	1	27	28	2.3
PHED	11	12	23	1.9
PHYS	15	39	54	4.4
Total	463	753	1220	100

Note: AEE: Aerospace Engineering; BIO: Biology; CE: Civil Engineering; CEIT: Computer Education and Instructional Technology; CENG: Computer Engineering; CHE: Chemical Engineering;
CHED: Chemistry Education; CHEM: Chemistry; EE: Electrical and Electronics Engineering; ENVE: Environmental Engineering; ESE: Elementary Science Education; FDE: Food Engineering; GEN: Genetics; GEO: Geological Engineering; IE: Industrial Engineering; MATH: Mathematics; ME: Mechanical Engineering; METE: Metallurgical and Materials Engineering; MINE: Mining Engineering; PETE: Petroleum and Natural Gas Engineering; PHED: Physics Education; PHYS: Physics. One subject in each * represented departments do not tick their gender; one subject does not tick his department. Therefore four missing presents in the demographic table but these subjects all included in the analysis.

Survey Adaptation Process

A rigorous translation process was followed instead of just translating the survey items. So instead of ‘translation’, the term ‘adaptation’ is preferred. Since survey adaptation includes all the activities from deciding whether a survey could measure the same construct in a different language and culture, to selecting translators, to adapting the test and checking its equivalence in the adapted form (Hambleton, 2005). Translation is only one of the steps in the process of survey adaptation. After getting permission from the first author of the CLASS, the CLASS in physics was firstly translated verbatim, and then adapted to Turkish language and culture by the researcher. Four experts in physics education reviewed the items in the CLASS in terms of two perspectives: (1) fitting the items to the categories proposed for the instrument, and (2) comparing the Turkish and English versions of the survey. One of the experts was the one who translated previously the MPEX into Turkish. Another expert did an independent English to Turkish translation and necessary modifications were made accordingly. An English teaching faculty checked the translation by comparing Turkish and English versions of the survey.

Student interviews were conducted with 13 students in think-aloud process. Each interview lasts for about 30 minutes. During the interviews, students were asked to rate the survey, and explain why they chose it and tell if any of the items in the survey was not clear. The researcher determined the items that were not clear to the participants and necessary modifications were done in consensus with the experts. During the adaptation process several discussions were made with Wendy Adams, one of the developers of the original CLASS survey, through e-mail. Among the items, Item 27 (*It is important for the government to approve new scientific ideas before they can be widely accepted*) was the one most of the interviewees reported to have confusion about its literal meaning. Similar problems with Item 27 were reported later in Zhang and Ding's (2013) study. In order to eliminate the confusion, some explanations were added to the beginning of the survey.

After the necessary modifications, researcher asked one Turkish language faculty to check the Turkish version of the survey in terms of clearness, punctuations, words etc., and the final form of the Turkish adaptation of the CLASS in physics was obtained. The Turkish translation of the CLASS has been available since than online[1] together with translations in other languages. Bayar and Karamustafaoglu (2015) administered the CLASS on 400 9th grade Turkish high school students; however, no research has been published yet about the findings of the obtained data from the CLASS administration in Turkish version over university students. In this study, survey adaptation and psychometric evaluation processes of CLASS in Turkish will be discussed for the interested researchers.

Data Analysis

For the statistical analysis of the data IBM Statistical Package for Social Sciences (SPSS 23.0) and IBM SPSS AMOS 23.0 programs were used. Firstly, a total of 18 negatively stated items were recoded (Items 1, 5, 6, 8, 10, 12, 13, 17, 18, 20, 21, 22, 23, 27, 29, 32, 35 and 40). Then, the data were screened out for the outliers and control item (Item 31). A missing data analysis was done for the preparation of data for further analysis. 171 subjects (12 %) were removed from further analysis of the study. The remaining data were randomly split into two for exploratory (n_{EFA}=610) and confirmatory factor analysis (n_{CFA}=610) for validity and model fit. The item and reliability analysis of the test scores were done.

Tests themselves cannot be valid or invalid. Instead, we validate the use of a test score in a specific context. Since validity is a matter of degree, not all or none, there exist no absolute method to determine the validity of the test scores. However, qualitative and quantitative evidence from different sources were collected to ensure validity. Validation, therefore, is the process of collecting evidence to support inferences (Crocker & Algina, 1986; Frankel & Wallen, 2000; Kane, 2006). For content-related evidence of validity, the items were examined and judged by the experts. Also, student interviews were done for face validity. For construct related evidence of validity factor analysis was used. There are two main approaches to factor analysis- exploratory and confirmatory. Exploratory factor analysis (EFA) is often used in the early stages of research to gather information about (explore) the interrelationships among a set of variables. It is primarily used to determine constructs for the measured items. EFA analysis performed in SPSS 23.0 by data reduction technique. In EFA, the correlations between all items were analyzed in order to select groups of items (also called as factors or categories) that all appear to measure the same idea. By examining the items that were grouped together, the researcher can determine if the survey measures the ideas it appears to measure. Principal component analysis (PCA), a form of factor analysis that is commonly used in scale development and evaluation, with varimax factor rotation was used. To check the factorability of the data Kaiser–Mayer–Olkin (KMO) measure of sampling adequacy value and Bartlett's test of sphericity were investigated. The data is suitable for factor analysis if the KMO value is above 0.60 and Bartlett's test of spherictiy should be statistically significant ($p< 0.05$) (Pallant, 2005). Tabachnick and Fidell (2015) suggested at least 300 cases for factor analysis. Pallant (2005) stated that the overall sample should be over 150 and there should be a ratio of at least five cases for each item to conduct factor analysis. The sample was 610 for EFA and 610 for CFA, and there are 41 items. Hence, the sample size was suitable for factor analysis. The other issue to be addressed concerns the strength of the inter-correlations among the items. Tabachnick and Fidell (2015) recommended an inspection of the correlation matrix for evidence of coefficients greater than 0.30.

Confirmatory factor analysis (CFA), on the other hand, is a more complex and sophisticated set of techniques used later in the research process to test (confirm) specific hypothesis or theories concerning the structure underlying a set of variables (Pallant, 2005). CFA analysis was performed on the other half of the data by using AMOS 23.0 to estimate how proposed models fit the data. When the value chi-square ($\chi2$) is significant, it is accepted that the data do not support the theoretical model. However, $\chi2$ is very sensitive to sample size (Kline, 2011). Therefore, standardized value of $\chi2$ and other model fit indices (model $\chi2$, goodness of fit test (GFI), adjusted goodness of fit index (AGFI), comparative fit index (CFI), normed fit index (NFI), non-normed fit index (NNFI), incremental fit index (IFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), parsimony normed fit index (PNFI), and parsimony goodness of fit index (PGFI)) were investigated. However, in order to evaluate how well a hypothesized model fit, acceptable statistics criteria and which model fit indices to be reported differ. Kline (2011) suggests the most important measures to report are the model chi-square, RMSEA, GFI, CFI, and SRMR. İlhan and Çetin (2014) summarized cut-off values for 'perfect fit' and 'acceptable fit' for the model statistics based on a rigorous literature review. In this study these summarized criteria were used. Table 4 gives these cut-off values together with obtained statistics in this study.

Results

The average scores for each item ranged from 2.29 to 3.99 for EFA data and from 2.31 to 3.97 for CFA data, with standard deviation values from 0.98 to 1.33 for EFA data and 0.94 to 1.36 for CFA data. Table 2 illustrates the mean, standard deviation, skewness, kurtosis, and item-total correlations for EFA and CFA data sets. Eleven items in EFA data set, and nine items in the CFA data set were found to have item-total correlations below 0.20 which means they might measure something different than other items in the survey. Those items were carefully considered whether during exploratory factor analysis they were deleted or not. None of those items were included in EFA.

Table 2
Descriptive Statistics and Item-Total Correlations of the CLASS in EFA and CFA

	Exploratory Factor Analysis (n=610)					Confirmatory Factor Analysis (n=610)				
Item	**M**	**SD**	**Skewness**	**Kurtosis**	r_{corr}	**M**	**SD**	**Skewness**	**Kurtosis**	r_{corr}
I1	3.10	1.252	-.192	-1.104	.408	3.22	1.226	-.277	-.969	.324
I2	3.57	1.076	-1.104	1.186	.238	3.54	1.132	-1.073	.877	.217
I3	3.20	1.286	-.530	-.583	.468	3.23	1.249	-.415	-.819	.464
I4	3.99	1.065	-1.298	1.708	.143	3.97	1.039	-1.173	1.198	.122
I5	2.65	1.212	.205	-.895	.355	2.62	1.180	.231	-.911	.339
I6	3.77	1.188	-1.070	.510	.400	3.82	1.189	-1.083	.606	.293
I7	3.17	1.260	-.506	-.047	.026	3.19	1.303	-.576	-.011	.037
I8	2.29	1.096	.661	.303	.049	2.31	1.050	.708	.380	.063
I9	3.26	1.219	-.540	-.364	.251	3.17	1.195	-.236	-.691	.336
I10	3.92	.983	-1.117	1.230	.281	3.90	1.045	-1.286	1.735	.253
I11	3.85	1.136	-1.029	.670	.354	3.89	1.162	-1.147	1.008	.381
I12	2.34	1.318	.517	-.878	.168	2.36	1.291	.591	-.783	.083
I13	3.30	1.299	-.459	-.797	.408	3.23	1.279	-.413	-.844	.334
I14	2.57	1.179	.141	-.759	.317	2.60	1.221	.190	-.774	.297
I15	3.56	1.030	-1.026	1.085	.359	3.56	1.057	-.882	.778	.348
I16	2.83	1.260	.021	-.979	.197	2.83	1.280	-.096	-.958	.269
I17	2.81	1.216	-.370	.041	-.066	2.76	1.200	-.319	-.119	-.090
I18	3.21	1.240	-.457	-.449	.174	3.28	1.189	-.468	-.388	.182
I19	3.54	1.133	-1.126	.969	.287	3.40	1.175	-.827	.013	.246
I20	3.12	1.245	-.275	-.901	.413	3.07	1.244	-.167	-1.070	.432
I21	3.44	1.308	-.714	-.329	.422	3.57	1.219	-.855	.000	.382
I22	2.70	1.066	-.010	-.281	.163	2.78	1.075	-.042	-.529	.236
I23	3.58	1.110	-1.121	1.151	.199	3.63	1.062	-1.111	1.295	.203
I24	3.95	1.085	-1.435	2.216	.403	3.95	1.076	-1.328	1.635	.359
I25	3.05	1.332	-.296	-.952	.505	2.98	1.324	-.221	-1.045	.554
I26	3.66	1.139	-1.560	2.647	.362	3.70	1.077	-1.515	2.842	.389
I27	3.91	1.304	-1.287	1.085	.177	3.89	1.360	-1.274	.856	.229
I28	3.44	1.134	-.779	.244	.482	3.59	1.098	-.868	.598	.480
I29	3.78	1.225	-1.067	.537	.493	3.74	1.209	-1.021	.439	.421
I30	3.51	1.187	-.914	.412	.552	3.54	1.066	-.772	.283	.508
I32	3.82	1.135	-1.051	.495	.463	3.86	1.171	-1.132	.743	.490
I33	3.24	1.127	-.576	-.031	.128	3.31	1.075	-.444	-.190	.058
I34	3.25	1.042	-.882	.838	.416	3.28	1.005	-.681	.515	.438
I35	3.52	1.194	-.790	.073	.475	3.52	1.198	-.657	-.313	.405
I36	3.39	1.131	-.713	-.149	.419	3.25	1.198	-.611	-.345	.414
I37	3.30	1.117	-.689	.046	.459	3.30	1.151	-.639	-.130	.380
I38	3.04	1.191	-.234	-.712	.149	3.03	1.245	-.350	-.727	.109
I39	3.69	1.024	-1.384	2.248	.347	3.78	.943	-1.256	2.196	.377
I40	3.50	1.191	-.815	.365	.506	3.47	1.092	-.564	-.180	.555
I41	3.13	1.200	-.544	-.110	-.055	3.17	1.235	-.438	-.419	.018
I42	3.57	1.111	-1.119	1.227	.426	3.66	1.124	-1.174	1.305	.430

Note: M: Mean, SD: Standard Deviation, r_{corr}: Item-total correlations

Firstly, an EFA was performed with all 41 items (except control item, I31) in the CLASS, but not a meaningful loading of items to the factors obtained. Secondly, EFA was performed with 26 items in the proposed model of Adams and his collogues (2006). Items loaded to six factors. In order to reduce distortions an iterative removal of items with communalities less than 0.30 was done (Tabachnick & Fidell, 2015). As a result items 8, 11, 16, 21, 22, and 23 were removed and analysis repeated with the remaining 20 items. The KMO value was found to be 0.888 and Bartlett's test was found to be statistically significant (p=0.00). Therefore, it could be concluded that data were suitable for the factor analysis. Following this finding, and as a result of the EFA principal components factorization technique and varimax rotation method, a three-factor model consisting of 20 items that explained 39.61 % of the total variance was obtained.

The obtained model in EFA was comprised of three factors. First factor includes seven items pertaining to the student problem solving and effort in solving physics problems. This factor is called *Problem solving effort.* Second factor includes six items pertaining to the student conceptual understanding and conceptual connections in physics. This factor is called *Conceptual understanding*. The last factor includes seven items pertaining to student internalization physics concepts and relating them to the real world around them. This factor is called as *Personal interest and real world connection*. Table 3 shows the factor loadings of the items in the final form of the scale.

Cronbach's alpha was calculated as a measure of internal consistency for each factor separately and for all the scale including 20 items for EFA data. According to reliability analysis, *Problem solving effort* $\alpha = 0.68$; *Conceptual understanding* $\alpha = 0.657$; *Personal interest and real world connection* $\alpha = 0.766$; and Overall scale $\alpha = 0.853$ were found.

The CFA was used to check the EFA results and the measurement model that was theoretically constructed. The 20-item model obtained from EFA was analyzed through confirmatory factor analysis (CFA) in AMOS 23. Based on modification indices, some error covariances were added to increase model fit indices. Table 4 gives the CFA model fit indices of this research together with acceptable reference fit indices. CLASS's fit index values obtained from the analysis were: χ_2/df=2.410, GFI=0.940, AGFI=0.923, CFI=0.901, NFI=0.843, NNFI=0.883, IFI=0.902, RMSEA=0.048, SRMR= 0.045, PNFI=0.719, and PGFI=0.725. Results from the CFA model provided evidence that three-factor model proposed in this study showed model fit, except for the NFI and NNFI.

The appropriate way to assess the fit of CFA models has been a subject of debate since 1970s. A plethora of fit statistics has been developed and discussed in the literature. What should be reported in CFA is not universally agreed upon. Although there is no universally agreed upon number of fit indices to report, a minimal set would include the χ^2 value and the associated degrees of freedom and a total of four approximate fit indexes that are among the most widely reported in the literature; RMSEA, GFI, CFI (Tabachnick & Fidell, 2015), and SRMR (Kline, 2011). These values are shaded in Table 4 and all in 'acceptable fit' or 'perfect fit' regions for the model proposed in this study.

Table 3

Factor Loadings of the CLASS as a Result of Explanatory Factor Analysis (EFA)

	Item statement	Problem solving effort (Factor1)	Conceptual understanding (Factor 2)	Personal interest & Real world connection (Factor 3)
I39	When I solve a physics problem, I explicitly think about which physics ideas apply to the problem.	.686		
I36	There are times I solve a physics problem more than one way to help my understanding.	.621		
I15	If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works.	.492		
I26	In physics, mathematical formulas express meaningful relationships among measurable quantities.	.462		
I42	When studying physics, I relate the important information to what I already know rather than just memorizing it the way it is presented.	.460		
I24	In physics, it is important for me to make sense out of formulas before I can use them correctly.	.459		
I34	I can usually figure out a way to solve physics problems.	.454		
I1	A significant problem in learning physics is being able to memorize all the information I need to know.		.674	
I6	Knowledge in physics consists of many disconnected topics.		.587	
I32	Spending a lot of time understanding where formulas come from is a waste of time.		.578	
I40	If I get stuck on a physics problem, there is no chance I'll figure it out on my own.		.544	
I5	After I study a topic in physics and feel that I understand it, I have difficulty solving problems on the same topic.		.533	
I13	I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.		.491	
I14	I study physics to learn knowledge that will be useful in my life outside of school.			.739
I3	I think about the physics I experience in everyday life.			.696
I30	Reasoning skills used to understand physics can be helpful to me in my everyday life.			.551
I37	To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.			.525
I25	I enjoy solving physics problems.			.502
I28	Learning physics changes my ideas about how the world works.			.501
I35	The subject of physics has little relation to what I experience in the real world.			.356

Table 4

Acceptable and Perfect Fit Values and Fit Indices Obtained in CFA

Fit indices examined	Criteria for perfect fit	Criteria for Acceptable fit	Fit Indices Obtained	Decision
χ^2/df	$0 \le \chi^2/df \le 2$	$2 \le \chi^2/df \le 4$	2.410	Acceptable fit
GFI	$0.95 \le GFI \le 1.00$	$0.90 \le GFI \le 0.95$	0.940	Acceptable fit
AGFI	$0.90 \le AGFI \le 1.00$	$0.85 \le AGFI \le 0.90$	0.923	Perfect fit
CFI	$0.95 \le CFI \le 1.00$	$0.90 \le CFI \le 0.95$	0.901	Acceptable fit
NFI	$0.95 \le NFI \le 1.00$	$0.90 \le NFI \le 0.95$	0.843	Not fit
NNFI	$0.95 \le NNFI \le 1.00$	$0.90 \le NNFI \le 0.95$	0.883	Not fit
IFI	$0.95 \le IFI \le 1.00$	$0.90 \le IFI \le 0.95$	0.902	Acceptable fit
RMSEA (90 % CI)	$0 \le RMSEA \le 0.05$	$0.05 \le RMSEA \le 0.08$	0.048	Perfect fit
SRMR	$0 \le SRMR \le 0.05$	$0.05 \le SRMR \le 0.10$	0.045	Perfect fit
PNFI	$0.95 \le PNFI \le 1.00$	$0.50 \le PNFI \le 0.95$	0.719	Acceptable fit
PGFI	$0.95 \le PGFI \le 1.00$	$0.50 \le PGFI \le 0.95$	0.725	Acceptable fit

Note: CI, confidence interval. $\chi^2 = 390.422$ $df = 162$ $p< 0.001$ (Table modified from İlhan & Çetin, 2014)

Standardized factor loadings with three-factor model obtained during the CFA are illustrated in Figure 1. Factor loadings ranged between 0.39 and 0.52 for F1 (*Problem solving effort*), 0.35 and 0.66 for F2 (*Conceptual understanding*), and 0.48 and 0.72 for F3 (*Personal interest & Real-world connection*). Table 5 compares the original version of CLASS categories to the proposed categories in the present study after data analysis.

Cronbach's alpha was calculated as a measure of internal consistency for each factor separately and for all the scale including 20 items for CFA data. According to reliability analysis, *Problem solving effort* α = 0.658; *Conceptual understanding* α = 0.644; *Personal interest and real world connection* α = 0.778; and Overall scale α=0.835 were found.

In order to compare the three-factor 15 items model proposed in Douglas et al. (2014), CFA was run over the present data set, since replicating a model in a study provides evidence of the trustworthiness of the original results. Model fit indices were firstly calculated for the 3-factor 15 items and correlated errors in Douglas et al. (2014)'s, and secondly for the 3-factor 14 items and correlated errors in Kontro & Buschhüter (2020)'s model. Table 6 compares model fit indices for the three models proposed in three research based on present data.

Figure 1

Standardized factor loadings for CFA model

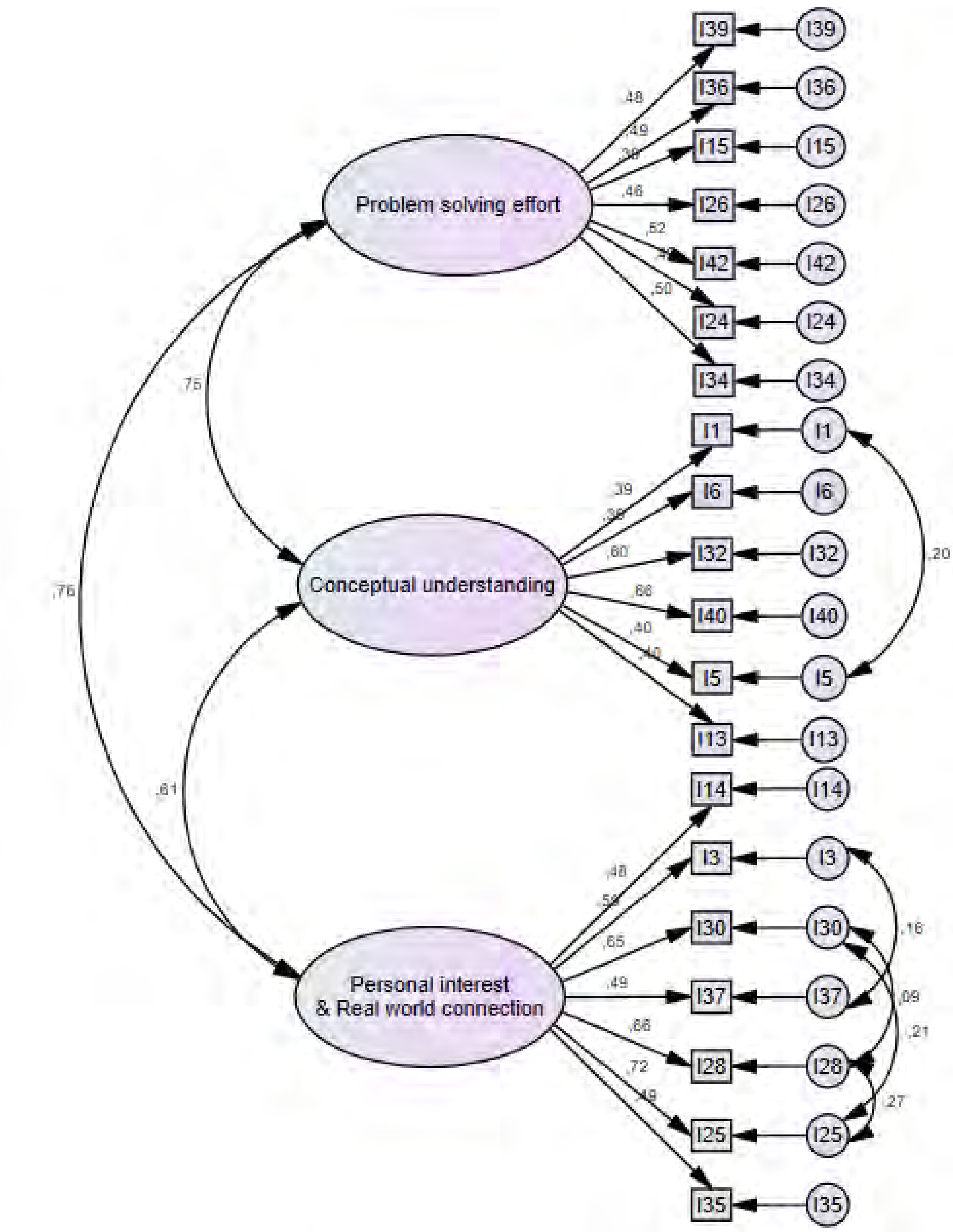

Table 5

CLASS Reported Categories by Adams et al. (2006), Proposed Categories in Douglas et al. (2014), Kontro & Buschhüter (2020) and in the Present Study

Original CLASS (Adams et al., 2014)		**Proposed CLASS in Douglas et al. (2014) and Kontro & Buschhüter (2020)**	
Categories	**Items**	**Categories**	**Items**
Real World Connection	**28, 30,** 35, 37	Personal Application and Relation to Real World	3, 14, 25*, 28, 30, 37
Personal Interest	**3, 11**, 14, **25**, 28, **30**	Problem Solving /Learning	5, 21, 22, 34, 40
Sense Making/Effort	**11**, 23, 24, **32,** 36, 39, **42**	Effort/Sense Making	23, 24, 29, 32
Conceptual Connections	**1**, **5**, **6**, **13**, **21**, **32**	**Proposed CLASS in Present Study**	
Applied Conceptual Understanding	**1**, **5**, **6**, 8, **21**, **22**, **40**	**Categories**	**Items**
Problem Solving General	**13**, **15**, **16**, **25**, 26, **34**, **40**, **42**	Problem Solving Effort	15, 24, 26, 34, 36, 39
Problem Solving Confidence	**15**, **16**, **34**, **40**	Conceptual Understanding	1, 5, 6, 13, 32, 40
Problem Solving Sophistication	**5**, **21**, **22**, **25**, **34**, **40**	Personal Interest & Real World Connection	3, 14, 25, 28, 30, 35, 37
Not Scored	4, 7, 9, 31, 33, 41		

Note: Items in bold are in more than one category. *In Kontro & Buschhüter (2020)'s model item 25 is excluded.

Discussion and Conclusions

The 42-item CLASS instrument was translated and adapted into Turkish to measure students' beliefs about physics and about learning physics. The Turkish version of the CLASS was given to a sample of 1331 freshman students enrolled in General Physics Laboratory II in a state university. Expert reviews and student interviews were used to ensure validity evidence to support the data. When the EFA was performed over all the CLASS items and over 26 items proposed in the original CLASS model, the former produced no meaningful model whereas the latter produced a three-factor model with 20 items. The CFA was performed over 20-item 3-factor model to evaluate how well the hypothesized model fit the acceptable statistics criteria. The model obtained in this study was compared in Table 5 and Table 6 with the models in the original CLASS and the model proposed by Douglas et al. (2014) and Kontro & Buschhüter (2020).

Table 6

Comparison of Model Fit Indices for the Model in Douglas et al. (2014), the Model in Kontro & Buschhüter (2020) and the Model in the Present Study

Fit indices examined	Douglas' et al. (2014) with 3 factor 15-item model	Decision	Kontro & Buschhüter's (2020) with 3 factor 14-item model	Decision	Present Study with 3 factor 20-item model	Decision
χ^2/df	3.047	AF	2.691	AF	2.410	AF
GFI	0.945	AF	0.958	PF*	0.940	AF
AGFI	0.921	PF*	0.937	PF*	0.923	PF*
CFI	0.902	AF	0.917	AF	0.901	AF
NFI	0.862	NF	0.876	NF	0.843	NF
NNFI	0.876	NF	0.892	NF	0.883	NF
IFI	0.903	AF	0.918	AF	0.902	AF
RMSEA (90% CI)	0.058	AF	0.053	AF	0.048	PF*
SRMR	0.051	AF	0.048	PF*	0.045	PF*
PNFI	0.681	AF	0.674	AF	0.719	AF
PGFI	0.654	AF	0.639	AF	0.725	AF

Note: AF: Acceptable fit, PF: Perfect fit, NF: Not fit. Model fit indices were calculated for the 3-factor 15-items and correlated errors in Douglas et al. (2014) and 3-factor 14-items and correlated errors in Kontro & Buschhüter (2020) model. * Represents the perfect fits.

Compared to the original CLASS, the proposed 20-item model was shorter, easier to administer and easier to score, valid and reliable survey, so feasible to use. The overlapping of the constructs (i.e. violation of the unidimensionality) in the original CLASS was a problem, while in the proposed model none of the items were included in more than one construct. The model-fit indices obtained from the CFA analysis were all in 'acceptable' or 'perfect fit' regions except for the normed fit index (NFI) and non-normed fit index (NNFI). The NFI evaluates the estimated model by comparing the χ^2value of the model to the χ^2 value of the independence model. High values (greater than .95) are indicative of a good- fitting model. However, the NFI may underestimate the fit of the model in good- fitting models with small samples. The adjusted NNFI improves on the problem of underestimating the fit in extremely good- fitting models but can also be much too small in small samples, indicating a poor fit when other indices indicate an adequate fit (Tabachnick & Fidell, 2015, p.721). The problem of the large variability in the NNFI is addressed by the IFI (Tabachnick & Fidell, 2015), which was at acceptable fit in the model. In addition to 20-item model proposed in the present study, the 3-factor 15-item model proposed by Douglas et al. (2014) and 3-factor 14-item model proposed by Kontro & Buschhüter (2020) were evaluated and the model-fit indices were compared in Table 6. In this way their model was empirically tested on a different sample. The NFI and NNFI values were also out of acceptable fit for these two models. The AGFI values indicated perfect fit in all three models. In addition to AGFI, in Kontro & Buschhüter's (2020) model GFI and SRMR values, and in the present study RMSEA and SRMR indices indicated perfect fit. Several similarities were found between the determined factors and items loading to these factors in those three models. Future research should conduct similar psychometric analysis to improve model fit, and to test those models in different samples. Since validity is a matter of degree and qualitative and quantitative evidence from different sources should be collected to ensure it.

CLASS items were purposefully worded broadly for the survey to be useful in several different contexts (Adams et al. (2006)) also made the responses fall into different interpretations. This issue was encountered in many items falling into different categories in the original CLASS. An important finding that emerged in all of these three studies (the present study, Douglas et al. (2014), and Kontro & Buschhüter (2020)), however, was that some of the eight factors proposed in the original CLASS should be combined. In the present study, some items in *Problem Solving General/Confidence* and *Sense Making/Effort* categories fell into same factor and called as *Problem Solving Effort*. When the items falling into this category investigated, they were found to be mostly related with effort in problem solving in physics such as 'If I get stuck on a physics problem my first try, I usually try to figure out a different way that works' (item 15) and 'I can usually figure out a way to solve physics problems' (item 34) which were originally in *Problem Solving Confidence*, *Problem Solving General* and *Problem Solving Sophistication* categories. Five of the six items (except item 21) loading into *Conceptual Understanding* category were in the *Conceptual Connections* category in the original CLASS. Four of the items loading into the same category were in the *Applied Conceptual Understanding* in the original CLASS. Six items falling into the same factor in the present study were investigated and named as *Conceptual Understanding*. Five of the six items (except item 11) in the original *Personal Interest* category and all of the four items in the original *Real World Connection* category fell into the same factor in the present study. Hence, the new category called as *Personal Interest and Real World Connection*.

Up to date, the CLASS has been used for different purposes including making distinctions between beliefs of experts compared to beliefs of novices; investigating the changing attitudes from the beginning to the end of a semester; researching the differences in attitudes and beliefs according to gender, country or teaching methods used; or the relationship between attitudes and achievement. In their usage the CLASS categories were as in the original overlapping categories. The overlapping of the constructs in the original CLASS was considered as a problem. It makes it difficult to discriminate and interpret the results obtained because of the highly correlating categories. In the proposed model, on the other hand, none of the items were included in more than one construct. The physics education research community will have greater confidence in their research results if they used this shorter and valid version of the CLASS proposed in this study to determine three constructs, namely Problem Solving Effort, Conceptual Understanding, and Personal Interest and Real-World Connection.

The overall internal reliability ($\alpha = 0.853$ for EFA data set, and $\alpha = 0.835$ for CFA data set) of the three-factor 20-item model proposed in the study was high, but the Cronbach's alpha values were below 0.70 threshold for *Problem solving effort* and *Conceptual understanding* subscales in both EFA and CFA data sets. Similar findings were obtained in their subscales in Douglas et al. (2014), Kontro & Buschhüter (2020) and Heredia and Lewis (2012) as well as in the original CLASS (Adams et al., 2006). If internal consistency is low, then the content of the items may be heterogeneous. Future research should extend on this topic as well.

Acknowledgements

I would like to present special thanks to Professor Dr. Ali Eryılmaz for his valuable contribution during the survey adaptation process. I also thank to the experts who contribute to the study and instructors who allowed me to conduct the study in their classes.

References

Adams, W., Perkins, K., Podolefsky, N., Dubson, M., Finkelstein, N., & Wieman, C. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics- Physics Education Research, 52*(1), 1-14.

Alhadlaq, H., Alshaya, F., Alabdulkereem, S., Perkins, K. K., Adams, W.K., & Wieman, C. E. (2009). Measuring students' beliefs about physics in Saudi Arabia. *AIP Conference Proceedings 1179,* 69-72.

Bates, S. P., Galloway, R. K., Lopston, C., & Slaughter, K. A. (2011). How attitudes and beliefs about physics change from high school to faculty. *Physical Review Special Topics- Physics Education Research, 7, 020114,* 1-8.

Bayar, A. & Karamustafaoglu, O. (2015). The Colorado Learning Attitudes about Science Survey (CLASS): The Study of Validity and Reliability. *International Journal of Assessment Tools in Education, 2*(1), 40-57.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Philadelphia: Holt, Rinehart and Winston, Inc.

Douglas, K. A., Yale, M. S., Bennett, D. E., Haugan, M. P., & Bryan, L. A. (2014). Evaluation of Colorado Learning Attitude about Science Survey. *Physical Review Special Topics- Physics Education Research, 10,* 020128, 1-10.

Downing, S. M. (2006). Twelve steps for effective test development. In S.M. Downing, & T. M. Haladayna (Eds.), *Handbook of test development* (pp. 3-25), New Jersey: Lawrence Erlbaum Associates, Inc.

Elby, A., Fredriksen, J., Schwarz, C., White, B. (2006). The Epistemological Beliefs Assessment About Physical Science (EBAPS), http://www2.physics.umd.edu/~elby/EBAPS/EBAPS_items.htm

Frankel, J. R., & Wallen, N. E. (2000). *How to design and evaluate research in education* (4th ed.). US: McGraw-Hill Comp.

Halloun, I. & Hestenes, D. (1998). Interpreting VASS dimensions and profiles for physics students. *Science Education, 7*, 553-.577.

Hambelton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* (pp.3-38), Matwah, NJ: Erlbaum.

Heredia, K., & Lewis, J. E. (2012). A psychometric evaluation of the Colorado Learning Attitudes about Science Survey for use in chemistry. *Journal of Chemical Education, 89*, 436-441.

İlhan, M. & Çetin, B. (2014). Development of clasroom assessment environment Scale (CAES):

Validity and reliability study. *Education and Science, 39*(176), 31-50.

Kane, M. (2006). Content-related validity evidence in test development. In S.M. Downing, & T. M. Haladayna (Eds.), *Handbook of test development* (pp. 131-154), New Jersey: Lawrence Erlbaum Associates, Inc.

Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd Ed.). NY: The Guilford Press.

Kontro, I., & Buschhüter, D. (2020). Validity of Colorado Learning Attitudes about Science Survey for a high-achieving, Finnish population. *Physical Review Physics Education Research, 16*, 020104, 1-11.

Kost, L. E., Pollock, S., & Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Physical Review Special Topics- Physics Education Research, 5, 010101, 1-14.*

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwatz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497-501.

Lindsey, B. A., Hsu, L., Sadaghiani, H., Taylor, J. W., & Cummings, K. (2012). Positive attitudinal, shifts with Physics by Inquiry curriculum across multiple implementations. *Physical Review Special Topics- Physics Education Research, 8*, 010102, 1-8.

Linn, R. L. (2006). The standards for educational and psychological testing: guidance in test development. In S.M. Downing, & T. M. Haladayna (Eds.), *Handbook of test development* (pp. 27-38), New Jersey: Lawrence Erlbaum Associates, Inc.

Marusic, M. & Slisko, J. (2012). Effects of two types of physics learning on the results of CLASS test. *Physical Review Special Topics- Physics Education Research, 8, 010107*- 1-12.

McMillan, J. H. & Schumacher, S. (2001). *Research in education: a conceptual introduction*. NY: Longman.

Milner-Bolotin, M., Antimirova, T., Noack, A., & Petrov, A. (2011). Attitudes about science and conceptual physics learning in university introductory physics courses. *Physical Review Special Topics- Physics Education Research, 7*, 020107, 1-9.

Pallant, J. (2005). *SPSS survival manual*. (2nd Ed.). NY: Open University Press.

Pawl, A., Barrantes, A. & Pritchard, D. E. (2012). What do seniors remember from freshman physics? *Physical Review Special Topics- Physics Education Research, 8, 020118, 1-12.*

Perkins, K. K., Adams, W. K., Pollock, S. J., Finkelstein, N. D., & Wieman, C. E. (2004). Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey, *in Proceedings of the Physics Education Research Conference, Sacramento, California*.

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American Journal of Physics, 66*(3), 212-224.

Sawtelle, V., Brewe, E., & Kramer, L. (2009). Validation study of the Colorado Learning Attitudes about Science Survey at a Hispanic-serving institution. *Physical Review Special Topics- Physics Education Research, 5,* 023101, 1-5.

Tabachnick, B.G. & Fidell, L. S. (2015). *Using multivariate statistics*. (6th Ed.). NY:HarperCollins.

Zhang, P. & Ding, L. (2013). Large-scale survey of Chinese precollege students' epistemological beliefs about physics: A progression or a regression? *Physical Review Special Topics- Physics Education Research, 9,* 010110, 1-9.

[1] See https://www.physport.org/assessments/assessment.cfm?A=CLASS for the Turkish translation of the CLASS.

[2] See https://www.colorado.edu/sei/class for the CLASS survey versions in physics, chemistry and biology.

Corresponding Author Contact Information:

Author name: Derya Kaltakci-Gurel

Department: Education Faculty

Faculty: Mathematics and Science Education Department

University, Country: Kocaeli University, Turkey

Email: deryakaltakci@gmail.com

Please Cite: Kaltakci-Gurel, D. (2021). Turkish Adaptation and Psychometric Evaluation of the Colorado Learning Attitudes About Science Survey (CLASS) in Physics. *The European Educational Researcher, 4*(3), 355 -372 https://doi.org/10.31757/euer.435

Received: July 25, 2021 ▪ Accepted: October 7, 2021