Assessing the Attitudes Towards Computing Scale: A Survey Validation Study

Dana Wanzer
The Findings Group
2646 Woodridge Drive
Decatur, Georgia USA
dana@thefindingsgroup.org

Tom McKlin
The Findings Group
2646 Woodridge Drive
Decatur, Georgia USA
tom@thefindingsgroup.org

Doug Edwards
Center for Education Integrating
Science Mathematics and Computing
Georgia Institute of Technology
Atlanta, Georgia USA
doug.edwards@ceismc.gatech.edu

Jason Freeman
Center for Music Technology
Georgia Institute of Technology
840 McMillan Street
Atlanta, Georgia USA
jason.freeman@gatech.edu

Brian Magerko
Digital Media Program
Georgia Institute of Technology
85 Fifth Street NW
Atlanta, Georgia USA
magerko@gatech.edu

ABSTRACT

Students who have positive attitudes towards computing are more likely to intend to persist in computing and ultimately persist in computing. Thus, this study describes the development and assessment of the Attitudes Towards Computing Scale, which was developed based on Williams et al. [20] Computer Science Attitude Survey. Assessment of the survey involved testing its reliability, dimensionality, and validity. Internal consistency reliability of the subscales and the total scale were strong. However, dimensionality tests using factor analysis did not support a five-factor structure; rather, the factor analyses suggested either using one item per subscale or mean composites of the subscales to form a total score of attitudes towards computing. This suggests there may not be meaningful differences among subscales. Tests of criterion validity show that the short-form of the scale, using a mean composite score of either one item per subscale or the composites of the subscales themselves, predict students' computing knowledge and intentions to persist just as well as using the subscales. Furthermore, an example is shown why using the subscales may be problematic due to multicollinearity issues. Recommendations for improving scales, such as not using reverse-coded items and having a sufficient number of items that differentiate between subscales, are provided. Ultimately, this study provides the computer science field with a scale of attitudes towards computing that demonstrates moderate validity and offers an example of how to assess other scales used in computer science education.

This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

SIGCSE '19, February 27–March 2, 2019, Minneapolis, MN, USA

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-5890-3/19/02.

<https://doi.org/10.1145/3287324.3287369>

KEYWORDS

Computing Attitudes; Survey Development; Survey Validation; Reliability; Validity; Factor Analysis

ACM Reference format:

D. Wanzer, T. McKlin, B. Magerko, D. Edwards, and J. Freeman. 2019. Assessing the Attitudes Towards Computing Scale: A Survey Validation Study. In *Proceedings of ACM SIGCSE Technical Symposium, Minneapolis, Minnesota, USA (SIGCSE 2019)*, ACM, NY, NY, USA, 7 pages. <https://doi.org/10.1145/3287324.3287369>

1 Introduction

Many researchers in the computer science community are focusing on ways to increase persistence in computing that will broaden participation in the field and help meet a growing demand for computing skills in the workforce. According to the theory of planned behavior, the most important precursor to behavior, such as persistence in the workforce, is behavioral intention, such as intentions to persist [1, 2]. There are three determinants of intentions to persist: perceived control over the behavior, subjective norms associated with the behavior, and attitudes towards the behavior (see Figure 1).

Figure 1. Theory of Planned Behavior for Predicting Persistence in Computing

Research on students' attitudes towards computer science suggest that some students, particularly female students, may have neutral or negative attitudes towards computing, which can detract from their likelihood of pursuing computing as a major or career [6, 17]. Thus, computer science education programs that target student attitudes towards computing are likely to improve

students' likelihood of persisting in computing. First, an adequate measure of attitudes towards computer science is needed.

Over a decade ago, Williams and colleagues [20] developed the Computer Science Attitude Survey (CSAS) to measure attitudes towards computer programming and computer science in general. This instrument was derived from the Fennema-Sherman Mathematics Attitude Scale [5], which was based on the premise that students' attitudes affect the way they view, pursue, and achieve within a given field.

The CSAS includes 57 items across five dimensions: (a) Confidence in Learning Computer Science and Programming, (b) Attitudes Towards Success in Computer Science, (c) Computer Science as a Male Domain, (d) Usefulness of Computer Science and Programming, and (e) Effective Motivation in Computer Science. The reliability of the subscales of the CSAS ranged from $\alpha = .82$ to $.91$ [20].

Researchers and evaluators have adapted the CSAS to respond to shifts in computer science education and the immediate needs of programs they work with. Researchers working with the EarSketch program have adapted the CSAS to form the *Attitudes Towards Computing* scale of the EarSketch Student Survey. The EarSketch Student Survey also includes a measure of intentions to persist and measures of creativity at both the person level and place level [10]. The Attitudes Towards Computing scale of the EarSketch Student Survey is based on the CSAS but is much shorter in length, consisting of only 19 items instead of 57. It does not include the *Computer Science as a Male Domain* and *Attitudes Towards Success* dimensions of the CSAS; rather, it includes two scales of *Computing Enjoyment* and *Computing Identity and Belongingness* which were created by the researchers. The subscales and items are shown in Table 1.

To date, no research has been conducted on the validity of the Attitudes Towards Success scale of the EarSketch Student Survey. Thus, this study examines the validity of the Attitudes Towards Success scale by examining its reliability, dimensionality, and content validity, which are commonly used techniques for assessing scale validity [18]. This study also aims to provide computer science researchers with an example of survey validation techniques that can be used with other scales in computer science.

2 Methods

2.1 Samples

Two samples of high school students were used in this study. Both samples came from a larger research and evaluation study on EarSketch [10]. The first sample consisted of students participating in the 2016-17 year of EarSketch ($n = 439$) and the second sample consisted of students participating in the 2017-18 year of EarSketch ($n = 423$). The 2016-17 year of EarSketch was a quasi-experimental design and included treatment, quasi-comparison, and comparison group students; for the purposes of this study, we examine all students who participated in the study and not just those who were in the treatment group because we believe the

construct would have similar reliability and dimensionality across all students.

Table 1. The Attitudes Towards Computing scale items

Computing Confidence
1. I have a lot of self-confidence when it comes to computing.
2. I am sure I could do advanced work in computing.
3. I think I could handle difficult computing problems.
4. I'm not good at computing. (R)
Computing Enjoyment
1. I enjoy computing.
2. Computing makes me nervous. (R)
3. Computing is difficult. (R)
4. I feel comfortable computing.
Computing Importance and Perceived Usefulness
1. I will be able to get a good job if I learn computing.
2. I will use computing in many ways throughout my life.
3. Computing is of no relevance to my life. (R)
4. Taking computing classes is a waste of time. (R)
Motivation to Succeed in Computing
1. I like solving computing problems.
2. When a computing problem arises that I can't immediately solve, I stick with it until I have the solution.
3. When I am working on a computing problem that I can't immediately understand, I want to work harder to get it.
4. Figuring out computing problems does not interest me. (R)
Computing Identity and Belongingness
1. I feel like I "belong" in computer science.
2. I consider myself as a scientist, technologist, engineer, or mathematician.
3. I take pride in my computing abilities.

Note: Items marked with an (R) are reverse-coded items.

2.2 Procedures and Measures

At the beginning and the end of the EarSketch program, students took a test of their computing knowledge and the EarSketch Student Survey. Items on the EarSketch Student Survey were rated on a 5-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*, with a 3 = *neutral*.

3 Results: Reliability

Reliability was measured with internal consistency, which is the general agreement between multiple items that comprise one larger construct. Internal consistency reliability was conducted using the Omega Total instead of Cronbach's alpha because many of the assumptions of Cronbach's alpha like tau equivalence were violated (see [11] for more information on why Cronbach's alpha may not be an acceptable measure of internal consistency). Omega Total is conceptually related to Cronbach's alpha and interpreted in much the same fashion, though it is slightly more conservative in its estimates. Coefficients range in value from 0 to 1; the higher the score, the more reliable the generated scale. Nunally [13] indicates that between .65-70 is an acceptable reliability coefficient. Omega Total was calculated using the ci.reliability function in the MBESS package [8].

The Omega Totals for the five constructs and the overall mean composite of the Attitudes Towards Computing Scale ranged from .648 to .931 (see Table 2). Reliability was lowest for the enjoyment sub-scale, indicating that this sub-scale may need to be supplemented by other items or items may need to be improved. Not surprisingly, the Omega Total for the total scale outperformed the reliability for each of the sub-scales; this is likely due to the number of items in the total scale ($n = 19$) versus the number of items in each of the sub-scales ($n = 3$ or 4 items per subscale).

Table 2. Omega Total (ω) for Both Samples Across Sub-Scales and the Total Scale

Sample Reliability Measure	1		2	
	ω	r	ω	r
Confidence	.815	.58	.83	.4
			4	7
Enjoyment	.66	.5	.64	.4
	7	4	8	2
Importance	.807	.5	.775	.5
		3		4
Motivation	.822	.6	.79	.58
		1	6	
Identity	.815	.7	.822	.6
		3		2
Total	.93	.7	.92	.6
	1	3	7	2

Note: Omega Total (ω) and inter-item correlations (r) are shown.

4 Results: Dimensionality

Dimensionality refers to how many dimensions or sub-constructs there are to the data. It is measured using factor analysis techniques. Dimensionality was first explored using exploratory factor analysis (EFA) with the first sample and then confirmed with confirmatory factor analysis (CFA) using the second sample. Recommended resources for factor analysis include [12, 19], both of which have guides for conducting analyses in SPSS.

4.1 Exploratory Factor Analysis

EFAs were performed with the psych package [14] in R to determine how many dimensions or factors comprised the Attitudes Towards Computing Scale and how the items aligned to each factor. EFAs were performed using maximum likelihood estimation with an oblique (direct oblimin) rotation because subscales in the Attitudes Towards Computing Scale were likely to be correlated with one another. Initial tests to determine whether EFAs could be performed (i.e., Bartlett's test, Kaiser-Meyer-Olkin [KMO] test of sampling adequacy, and determinant of the matrix) were all satisfactory.

EFAs were evaluated based on extracted communalities and factor loadings [3]. Extracted communalities were assessed to see whether items in a factor are sufficiently related to one another; guidelines suggest that extracted communalities should be above .30. Factor loadings determine how strongly items are related to that factor; guidelines suggest that items should be at minimum above .30 but it is desirable that they are above .50. Furthermore,

items should only have factor loadings above .30 on one factor, otherwise they are crossloading across multiple factors and should be removed.

4.1.1 EFA 1: Two Factors with 19 Items. Parallel analysis was performed with the hornpa package [7] to determine how many factors to retain; results indicated retaining two factors, which does not align with the theoretical model suggesting five factors. Thus, a two-factor EFA was performed with all 19 items. Results were not favorable (see Table 3): one item did not load onto any factor greater than .30, two items were crossloading, and three items had extracted communalities less than .30 suggesting either the item is not related to the other items or there needs to be another factor. The two factors explained a total 48% of the variance of all the items, indicating that the set of items explains a large percentage of the variance in the latent variable of attitudes towards computing. Although the crossloading items are problematic, we first removed the three items with low extracted communalities to see if that is enough to improve the model.

Table 3. Results from EFA 1– Two Factors with All 19 Items

Item	F1	F2	EC	Issues
Confidence2	.856		.623	
Confidence1	.835		.632	
Confidence3	.822		.606	
Identity3	.632		.583	
Enjoyment4	.595		.609	
Identity1	.584		.648	
Identity2	.482		.392	
Enjoyment1	.456	.423	.629	Crossloading
Motivation1	.443	.430	.619	Crossloading
Enjoyment3_R	.395			Low EC
Confidence4_R	.385		.336	
Importance4_R		.806	.598	
Importance3_R		.784	.520	
Importance2		.569	.520	
Motivation4_R		.526	.497	
Motivation2		.443	.424	
Importance1		.418		Low EC
Motivation3		.400	.398	
Enjoyment2_R				Low FL; Low EC

Note: Items correspond to the numbers and factors in Table 1; items with a “R” at the end are reverse-coded items. F1 and F2 refer to the factor loadings for factors 1 and 2, respectively. EC = extracted communality, FL = factor loading

4.1.2 EFA 2: One Factor with 16 Items. The three problematic items from EFA1 were removed and a new parallel analysis performed with the remaining 16 items, which indicated retaining one factor. Thus, a second EFA was performed using one factor with the 16 items, which resulted in an acceptable solution (see Table 4): all items had extracted communalities and factor loadings above .30 and the factor explained a total 48% of the total variance of all the items.

4.1.3 EFA 3: One Factor with Best Item Per Subscale. An alternative EFA was tested with just the best-performing item (i.e., the one with the highest factor loading in EFA2) from each of subscale of the Attitudes Towards Computing Scale. This EFA resulted in an acceptable solution (see Table 5): all items had extracted communalities above the minimum of .30, factor loadings all exceeded .60, and the factor explained a total 59% of the variance of all the items.

Table 4. Results from EFA 2 – One Factor with 16 Items

Item	Factor Loading	Extracted Communality
Identity1	.811	.658
Enjoyment1	.798	.636
Motivation1	.797	.635
Enjoyment 4	.775	.601
Identity3	.755	.570
Confidence1	.717	.514
Confidence3	.695	.483
Confidence2	.692	.479
Importance2	.683	.467
Motivation4_R	.680	.462
Motivation2	.633	.400
Motivation3	.629	.395
Identity2	.629	.395
Importance4_R	.619	.383
Confidence4_R	.576	.332
Importance3_R	.548	.300

4.1.4 EFA 4: One Factor with Subscale Composites. An alternative EFA was also tested with mean composites of each of the five subscales of the Attitudes Towards Computing Scale. This was performed because factor analyses often use parcels (i.e., means) of items to reduce the overall number of items analyzed and improve analyses [9]. This EFA resulted in an acceptable solution (see Table 5): all items had extracted communalities above the minimum of .30, factor loadings all exceeded .70, and the factor explained a total 64% of the variance of all the composite items.

4.1.5 EFA Conclusions. Overall, the theoretical 5-factor model of the Attitudes Towards Computing Scale was not supported; parallel analysis suggested only one or two factors underlying the data. Removing poorly performing items (see Table 3) improved the model, but the third and fourth EFAs—consisting of one item per subscale and the mean composites for each subscale, respectively—performed the best of all models with extracted communalities factor loadings far exceeding the minimum of .30. Ultimately, there may be too few items to result in five subscales or there may be a more dominant higher-order construct of attitudes towards computing, as evidenced in the 1-factor EFAs. These results also suggest that revisions of items may be needed as the subscales are not being perceived as sufficiently distinct by participants. However, although a longer scale may provide five clear subscales, it will also be more burdensome on participants.

Table 5. Results from EFAs 3 and 4

Construct	Best (EFA3)	Item	Composites (EFA4)	
	FL	EC	FL	EC
Identity	.826	.682	.851	.724
Motivation	.815	.664	.828	.686
Confidence	.813	.661	.819	.671
Enjoyment	.722	.521	.767	.588
Importance	.662	.438	.715	.511

4.2 Confirmatory Factor Analysis

CFAs were performed using maximum likelihood estimation using the lavaan package [15, 16] in R; each of the four models tested from the EFAs were tested again using CFA with the second sample. See Table 6 for the results of the CFAs.

The first CFA tested the theoretical model with all 19 items loading onto their respective subscale factor and the subscale factors loading onto a global computing attitudes factor. This model did not fit the data well, with none of the fit indices reaching goal values. The Beta factor loadings suggested the reverse-coded items had much lower factor loadings onto their respective factors compared to the non-reverse-coded items. Removing the reverse-coded items greatly improved the model fit (see Table 6, Model 1b). However, there were then too few items in two of the factors that had two reverse-coded items. Furthermore, the attitudinal factors were also extremely highly related to one another, with Beta factor loadings to the global attitudinal factor ranging from .815 to .927.

Table 6. CFA Results for All Four Models Tested

Model	X ² (df)	CFI	RMSEA	SRMR
Goal Values	Smaller	> .9	< .08	< .06
1: Theoretical Model	530.5 (147)	.889	.087	.065
1b: Theoretical minus reverse-coded items	137.5 (60)	.97 1	.061	.032
2: 16 items (EFA2)	555.4 (104)	.85 3	.112	.065
3: Best Item per Subscale (EFA3)	9.7 (5)	.99 4	.052	.017
4: Composites of subscales (EFA4)	20.3 (5)	.98 3	.094	.022

Note: X² = chi-square; df = degrees of freedom; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

The second CFA tested the results from EFA2 with all 16 items loading onto the general computing attitudes factor. This model did not fit the data well, with none of the fit indices reaching goal values. Examination of the modification indices suggested that similar items (e.g., Confidence2 and Confidence3, Importance3_R and Importance4_R, Motivation2 and Motivation3) were more highly correlated to one another than to the other items. Allowing these items to correlate, or removing one from

each pair of items, greatly improved model fit but suggests that participants are not perceiving these items as sufficiently different from one another.

Similar to the EFA results, the third and fourth models were the best performing CFA models; however, the third model consisting of only the best performing item from each subscale performed best with all fit indices exceeding goal values.

5 Results: Criterion Validity

Criterion validity, which assesses how well our survey predicts the outcomes of other scales (i.e., intent to persist in computing and computing knowledge), was analyzed with the both samples together. Mean composites were created based on the four models tested. For the first model, subscales were created based on the averages of each of the three or four items (see Table 1). For the second model, scores for all 16 items from EFA2 were averaged. For the third model, scores for the best item from each subscale (EFA3) were averaged. For the fourth model, scores for each sub-construct from the first model were averaged and then the average of those five sub-constructs was computed. Correlations among all the composite scores are shown in Table 7. Correlations among the subscale composites (top left box) range from $r = .51$ to $.71$, which are very high and would be problematic in regression analyses due to multicollinearity. Correlations among the three composites from the three EFA models (bottom right box) range from $r = .93$ to $.99$, suggesting scores across models are nearly identical.

Table 7. Correlation Among All Subscales and Total Scales

	1	2	3	4	5	6	7	8
1. Confidence								
2. Enjoyment	.6							
3.	.5	.5						
Importance	.6	1						
4. Motivation	.6	.6	.6					
5. Identity	.7	.6	.6	.70				
6. EFA2	.8	.7	.7	.88	.87			
7. EFA3	.78	.70	.7	.8	.85	.9		
8. EFA4	.85	.80	.7	.87	.88	.9	.9	

Note: Values in the top left box are the correlations among the subscales. Values in the bottom right box are correlations among the composites created from the three EFA models.

However, the mean composite of the best item from each subscale (EFA3) is not a perfectly continuous variable. Whereas EFA2 and 4 are more continuous (Figure 2a), EFA3 has small gaps between values across the x-axis (Figure 2b).

To test the criterion validity of the four models on outcome variables of interest, a series of regressions were performed such that the outcome variable was predicted by the mean composite(s)

from each model while controlling for retrospective pre-test scores of the outcome variable. For instance, the independent variables in the regression for model 1 consisted of the retrospective pre-test score of the outcome variable (e.g., Intent to Persist or computing knowledge at pretest) and the mean composites for each of the five subscales of the Attitudes Towards Computing Scale. This was repeated for each model across both outcome variables of interest. The R-squared values were compared to see if any model had greater predictive validity than other models.

Table 8 shows the resulting R-squared values. Overall, there were very minor differences in the criterion validity across models. This suggests that there is no real loss of information in the Attitudes Towards Computing scale when only a handful of items are used compared to using composites comprised of all or most items.

Figure 2. Correlations Between EFA 4 with EFAs 2 and 3

Table 8. R-Squared Values for Regressions Predicting Computing Knowledge and Intent to Persist

Model	Computing Knowledge	Intent to Persist
1: Five subscale composites	46.5%	71.8%
2: 16 items (EFA2)	45.7%	70.4%
3: Best Item per Subscale (EFA3)	45.2%	72.0%
4: Composite of the Composites of each subscale (EFA4)	45.8%	69.8%

5.1 Multicollinearity: An Example

When independent variables are strongly correlated, it may cause issues in statistical inference such that regression coefficients are unstable. To demonstrate how the strong correlations of the subscales of the Attitudes Towards Computing Scale can be problematic, another regression was performed predicting computing knowledge and intent to persist with retrospective pre-test scores of computing knowledge or intention to persist and all five subscales of the Attitudes Towards Computing Scale. Table 9 shows the simple correlations and regression coefficients for the first model examined in Table 8. Despite having moderate to strong positive correlations between each of the subscales and the two outcome variables, one of the standardized regression coefficients (Betas) was negative which is an example of suppression. The multicollinearity results in unstable Beta weights that fluctuate widely depending on which subscales are included in the regression. Therefore, it would be more advisable to use a mean composite of either one item per subscale (e.g., Model 3 in Table 8) or a mean composite of all the items (e.g., Model 2 or 4 in Table 8).

Table 9. Simple Correlations and Beta Coefficients for Computing Knowledge and Intent to Persist Post Scores

Subscale	Computing Knowledge		Intent to Persist	
	r	Beta	r	Beta
Pretest Score	.70		.597	
Confidence	.27	.007	.57	.021
Enjoyment	.24	.078	.54	.001
Motivation	.40	.135	.61	.121
Importance	.32	.028	.65	.174
Identity	.24	-.076	.72	.285

6 Conclusions

The purpose of this study was to examine the reliability, dimensionality, and validity of the Attitudes Towards Computing Scale. Overall, reliability of the subscales and the total scale of Attitudes Towards Computing was high. However, tests of dimensionality did not support the theorized 5-factor solution; rather, the items and subscales were so highly correlated that a 1-factor solution fit the data best. Both EFAs and CFAs supported using either one item per subscale or mean composites of each subscale. Furthermore, there was no loss of information in using a

total composite or composite of just one item per subscale when predicting Intent to Persist or Computing Knowledge.

7 Discussion

This research sheds light on some issues with common practices in survey development, including reverse-coded items, survey fatigue, and pretesting.

First, many scales contain reverse-coded items, but people interpret reverse-coded items differently than non-reverse-coded items, which can result in factor analysis issues [15]. Reverse-coded items tend to either not perform well with non-reverse-coded items or to load onto their own separate factor.

Second, researchers often have to balance having enough items to have meaningful subscales and having too many items that introduce survey fatigue in participants. Though the Attitudes Towards Computing scale did not provide meaningful subscales, we believe the total scale is still useful in measuring attitudes towards computing globally. This study resulted in both a long (16 items, or 13 items if reverse-coded items are removed) and short (5-items) form of the Attitudes Towards Computing scale.

Third, this study's failure to find support for the 5-factor solution highlights the importance of thoroughly pretesting surveys. Pretesting surveys can involve expert review, testing the survey through the Quality Appraisal System [21], and cognitive interview techniques [4]. If pretesting had been done with the items of the Attitudes Towards Computing Scale, many of the issues highlighted in this paper may have been identified prior to costly data collection efforts. Importantly, this study focused on criterion validity, but there are many other types of validity that were not discussed. For example, pretesting can be useful for construct validity (i.e., the extent to which a test measures the construct it claims it measures) and content validity (i.e., the extent to which a test covers the full range of the concept's meaning).

Ultimately, we believe this study highlights the need for researchers to rigorously test the scales they use. It is our hope that this study can serve as a guide for researchers in testing their own scales to determine whether they are reliable and valid for the populations they are using it with.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1504293. The curriculum described in this paper is available at <http://earsketch.gatech.edu/itec2120/book.html>.

REFERENCES

- [1] Ajzen, I. 2001. Nature and operation of attitudes. *Annual Review of Psychology*. 52, 1 (Feb. 2001), 27–58. DOI:<https://doi.org/10.1146/annurev.psych.52.1.27>.
- [2] Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 50, 2 (Dec. 1991), 179–211. DOI:[https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).
- [3] Costello, A.B. and Osborne, J.W. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*. 10, 7 (2005), 1–9.
- [4] De Leeuw, E., Borgers, N. and Smits, A. 2004. Pretesting questionnaires for children and adolescents. *Methods for testing and evaluating survey questionnaires*. S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer, eds. John Wiley & Sons, Inc. 409–429.

- [5] Fennema, E. and Sherman, J.A. 1976. Fennema-sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*. 7, 5 (Nov. 1976), 324. DOI:<https://doi.org/10.2307/748467>.
- [6] Funke, A., Berges, M. and Hubwieser, P. 2016. Different perceptions of computer science. *2016 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)* (Mumbai, India, Mar. 2016), 14–18.
- [7] Huang, F. 2015. *hornpa: Horn's (1965) Test to Determine the Number of Components/Factors*.
- [8] Kelley, K. 2007. Methods for the Behavioral, Educational, and Social Sciences: A R package. *Behavior Research Methods*. 39, 4 (Nov. 2007), 979–984. DOI:<https://doi.org/10.3758/BF03192993>.
- [9] Little, T.D., Cunningham, W.A., Shahar, G. and Widaman, K.F. 2002. To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*. 9, 2 (Apr. 2002), 151–173. DOI:https://doi.org/10.1207/S15328007SEM0902_1.
- [10] Magerko, B., Freeman, J., Mcklin, T., Reilly, M., Livingston, E., Mccoid, S. and Crews-Brown, A. 2016. EarSketch: A STEAM-based approach for underrepresented populations in high school computer science education. *Trans. Comput. Educ.* 16, 4 (Sep. 2016), 14:1–14:25. DOI:<https://doi.org/10.1145/2886418>.
- [11] McNeish, D. 2017. Thanks coefficient alpha, we'll take it from here. *Psychological Methods*. Online First, (May 2017). DOI:<https://doi.org/10.1037/met0000144>.
- [12] Meyers, L.S., Gamst, G. and Guarino, A.J. 2017. *Applied multivariate research: Design and interpretation*. SAGE.
- [13] Nunnally, J.C. 1978. *Psychometric theory*. McGraw-Hill.
- [14] Revelle, W. 2018. *psych: Procedures for Psychological, Psychometric, and Personality Research*.
- [15] Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*. 48, 2 (2012), 1–36. DOI:<https://doi.org/10.18637/jss.v048.i02>.
- [16] Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Chow, M. and Jorgensen, T.D. 2018. *lavaan: Latent Variable Analysis*.
- [17] Schulte, C. and Knobelsdorf, M. Attitudes towards computer science-computing experiences as a starting point and barrier to computer science. *Third International Workshop on Computing Education Research* (Atlanta, GA, USA).
- [18] Slavec, A. and Drnovšek, M. 2012. A perspective on scale development in entrepreneurship research. *Economic and Business Review*. 14, 1 (2012), 39–62.
- [19] Tabachnick, B.G. and Fidell, L.S. 2013. *Using multivariate statistics*. Pearson Education.
- [20] Williams, L., Wiebe, E., Yang, K., Ferzli, M. and Miller, C. 2002. In support of pair programming in the introductory computer science course. *Computer Science Education*. 12, 3 (Sep. 2002), 197–212. DOI:<https://doi.org/10.1076/csed.12.3.197.8618>.
- [21] Willis, G.B. and Lessler, J.T. 1999. Question Appraisal System - QAS-99. (1999), 84.