2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #37238

Cultivating an Inclusive Environment in Computer Science: Validity Evidence for a New Scale

Dr. Mary E Lockhart, Texas A&M University

Mary E. ”Betsy” Lockhart, Ph.D. is a Postdoctoral Research Associate at Texas A&M University. Her research focuses on what factors influence diverse students to choose and persist in STEM. Particularly, she is interested in the development and cultivation of students’ STEM identities and the potential protective element these identities have in student retention. Dr. Lockhart graduated from Stephen F. Austin State University with a B.S. in Mathematics and Psychology. She then graduated from Texas A&M University with a M.S. in Mathematics and Ph.D. in Educational Psychology with a specialization in Research, Measurement and Statistics.

Dr. Karen E Rambo-Hernandez, Texas A&M University

Karen E. Rambo-Hernandez is an associate professor at Texas A & M University in the College of Education and Human Development in the department of Teaching, Learning, and Culture. In her research, she is interested in the assessing STEM interventions on

Cultivating an Inclusive Environment in Computer Science: Validity Evidence for a New Scale

Abstract

This research paper describes the development and initial validation of an instrument to measure students' inclusive attitudes and behaviors within computer science and its sensitivity to intervention effects. The lack of diversity within computing degree programs and fields has been an ongoing concern for several years. National programs and initiatives have placed a high priority on broadening participation in computer science and making the computing culture more inclusive of women and ethnic groups typically underrepresented in computer science. Only small gains have been documented. In this study, we adapted and modified a measure of university students' valuing of diversity and willingness to act inclusively within engineering contexts—a field with similar diversity concerns—for computer science. Given computer science departments are often housed within engineering schools or colleges which often have similar problems with underrepresentation, adapting an engineering scale, specifically the Valuing Diversity and Enacting Inclusion in Engineering (VDEIE), for computer science (VDEI-CS) was considered both relevant and applicable. After adapting the scale, validation efforts utilizing confirmatory factor analysis and multiple indicators multiple causes structural equation modeling affirmed that the VDEI-CS produced valid and reliable scores to gauge measures of inclusion and diversity within computer science students. Next, a small pilot study was conducted to determine if the VDEI-CS was sensitive enough to detect the effect of a series of computer science interventions geared towards enhancing first-year students' value of diversity and willingness to act inclusively in computing contexts. Though the study was preliminary, results of the multilevel analysis demonstrated promise that the VDEI-CS was sensitive enough to detect changes as intervention groups demonstrated statistically significantly greater gains in their value of diversity to fulfill a greater purpose and willingness to act inclusively by promoting healthy behaviors than control groups. The primary purpose of this study is to relay to the computing research community a tool to assess student attitudes toward the value of diversity and inclusive behaviors in computing contexts that will enable researchers to gauge the temperature of a group of students and assess the effect of interventions developed to promote change within the culture.

Background

The need for computing professionals in the workforce is growing rapidly. The U.S. Bureau of Labor Statistics (2022) estimates that employment in computer and information technology occupations is projected to grow 15% from 2021 to 2031, generating 682,800 new jobs and annually replenishing another 418,500 vacancies. This rate is much faster than the average for all other occupations [1]. These statistics indicate that there is a great need to continue to increase the overall number of qualified computing professionals within the United States. Though the number of undergraduate students enrolled in computing majors has increased over the past decade and will hopefully continue to rapidly increase for the next ten years in order to meet the upcoming demand for new computing-filled professional vacancies, progress in creating a more diverse computing culture has been slow [2].

The computing culture has a reputation of lacking in diversity. Within the United States (U.S.), the disparities in computer science education and careers are glaring. Blacks, Hispanics, Native

Americans, Alaska Natives, and women are all underrepresented as compared to their relative proportions of the national population [2]. Of particular interest, though some of these groups have documented increases in their share of awarded computing degrees over the past decade, the percentage of computer science bachelor's degrees awarded to Blacks has actually decreased during this time [2]. Furthering the alarm, the percentage of computer science bachelor's degrees awarded to women is 8% lower than what it was two decades ago [2]. Although perfect proportionality is not the goal, the move toward less representation in computer science is concerning. Moreover, this lack of representation is not unique to the U.S. In the United Kingdom (U.K.), for example, the U.K. Department for Education documented that only 0.4% of women pursue computing degrees compared to 4.5% of men [3].

Diversity in the computing workforce is important. Building a diverse computing workforce broadens the computing research agenda and facilitates the construction of new, equity-centered, technologies [4]. Due to the power of diversity to foster creativity and provide new perspectives on a problem, diverse teams are more capable of solving challenging problems than teams comprised of members with similar levels of intelligence, but lacking in diversity [5], [6]. Yet, the computing culture continues to lack diversity and is, thus, hindered. Cultivating an inclusive culture within computing education contexts where diversity is valued is a potential solution. The Theory of Reasoned Action states that behaviors are a function of attitudes and perceived subjective norms [7]. Thus, computing students' likelihood to enact inclusive behaviors within computing contexts is a function of their attitudes toward diversity, inclusion, and equity in computing and their perceived norms of how relevant groups perceive diversity in computer science. To this end, we have adapted a scale used to measure engineering student's value of diversity and willingness to act inclusively in engineering contexts, a field with similar underrepresentation concerns, to computer science. The purposes of the studies presented here are to detail the assessment of this adapted scale, namely the Valuing Diversity and Enacting Inclusion in Computer Science (VDEI-CS), and to present this instrument to the research community for use in diversity investigations in computing.

Culture of Computing

Several explanations for the gaps in representation of individuals from typically underrepresented groups in computer science programs and careers have been suggested. Lack of access to computing technology, inadequate K-12 preparation, lack of role-models, stereotype threat, and lower self-efficacy have all been identified as reasons non-majority students do not enter or eventually leave computing programs [8]-[19]. Specifically in STEM fields and disciplines, non-majority students' sense of belonging is imperative to their retention and success within STEM programs and is associated with a variety of positive outcomes for individuals including: increased GPA, increased self-reported health and well-being, and increased academic scores [20], [21]. Yet, in direct opposition to non-majority students cultivating this sense of belonging, or fit, in computing is the reality of the computing education "culture" in the U.S. being primarily one-note (e.g., white-men)—including faculty, students, and professionals—which instigates perpetual curricular and non-curricular hurdles for members of non-majority groups to overcome. To attain their fit within computing, students must navigate the computer science culture by adopting norms and values that are reflective of the majority-group [22]. Not being able to adopt these norms and values impacts students' fit within computing contexts and, ultimately, their retention.

Culture is a compelling explanation for underrepresentation in computer science. This identified one-note cultural concern in computing contexts where non-majority computing students are continuously presented with faculty and student populations that lack in diversity creates barriers for a non-majority student to effectively integrate into the discipline [23]. Computer science faculty influence the maintenance and propitiation of the storyline regarding who belongs in computer science due to their "membership" status in the professional computing world and their belief of what constitutes a computer science professional [22], [24]. These faculty typically draw upon their own professional experiences and practices as majority members as a representation of "legitimate" work in computer science—including their research, educational background and curriculum. It follows that the faculty intentionally or unintentionally introduce norms and values into the computing culture based upon a majority-member reflection regarding who is or can be a computer scientist [22], [25]-[27]. Faculty, thus, play an important role in encouraging or discouraging majority member students to enact inclusive behaviors in computing as they pass along attitudes and subjective norms and values of the computing culture to these future computing professionals. If these norms and values do not purposefully value diversity and inclusion, then non-majority students face extra challenges to effectively integrate into the computing culture. Now these students are not only isolated by their non-majority status, but they must also adopt the norms and values that are reflective of majority members and lack in diversity values. If non-majority students cannot navigate through adopting these norms and values presented by the majority, they are found to lack in their internalization of their fit within the discipline and, ultimately, are at greater risk of leaving [27], [28].

The lack of diversity within the student population of computing programs is also a cultural concern and presents challenges to the retention of non-majority students within computing. For example, research has identified that classmates can be the most effective means of helping undergraduates cope with the difficulties of being computing majors [23], [29]. Yet, attaining this support system of classmates can be challenging for non-majority students as they simply lack access to other non-majority students. For example, Cohoon [29], identified that STEM departments with higher student proportions of women were more likely to retain women at comparable rates to men. Specifically, women computing students enrolled in programs who receive support from other women computing students are more likely to be retained to graduation than those who do not [23]. Unfortunately, it has also been realized that lower percentages of non-majority member representation in STEM disciplines contributes to a lesser sense of belonging and desire of these members to participate. For example, when women STEM majors were put in a group that was men-dominant, they reported a statistically significantly lower sense of belonging and desire to participate than women who were in a group with equal proportions of men and women STEM majors [30]. Given that non-majority computing students lack access to other computing students from similar underrepresented groups and that a support system of computing peers is of such great benefit for students, it is imperative that the majority members work to foster attitudes in computing settings that value diversity and are willing to act inclusively. This will contribute to non-majority students receiving the peer support they need and further enable them to internalize their fit within the discipline.

The VDEI-CS is an applicable tool to assess these computing students' attitudes towards the value of diversity and willingness to enact inclusive behaviors within computing contexts; thus, providing a gauge of the computing culture's value towards diversity, equity and inclusion practices and principles. The original VDEIE was developed by Rambo-Hernandez and colleagues [6] who noted there was a lack of instruments with validity evidence that would measure engineering students' value of diversity and willingness to act inclusively within engineering contexts—deemed their inclusive professional engineering identity. The VDEI-CS directly mirrors the original VDEIE and, thus, measures students' attitudes towards diversity and their willingness to act inclusively. This is done in conjunction with the original VDEIE by evaluating two central constructs—Valuing Diversity and Enacting Inclusive Behaviors (see Appendix A for a list of related survey items). Valuing Diversity is represented by the two primary factors of Serving Customers Better (VL-S) and Fulfilling a Greater Purpose (VL-F). A high score on VL-S would indicate that the computing student believed customers could be better served if diversity is valued. A high score on VL-F would indicate the computing student perceived valuing diversity aligned with a strong inward desire for purpose and fairness in their work. Furthermore, the Enacting Inclusive Behaviors construct is represented by the two primary factors of Promoting Healthy Behaviors (BH-P) and challenging discriminatory behaviors (BH-C). A high score on BH-P would indicate the engineering student would take measures to ensure every team member was included and valued and sought to have a variety of skills represented on the team. A high score on BH-C would indicate that the engineering student would call out any type of discriminatory behavior while working on a team. Taken together, the VDEI-CS would allow researchers to measure the proverbial temperature of a group of students within the computing culture and assess the effect of interventions developed to change the culture of computing. Given that the VDEI-CS could be of great importance to the computing culture, investigating the validity and reliability of the instrument is critical.

Present Study

Working towards cultural change and increasing diversity within computing contexts is a noteworthy, yet complicated endeavor. A variety of structural and curricular steps can be taken to cultivate a more diverse, equitable, and inclusive culture within computing educational settings and to provide the groundwork for graduates to carry these more inclusive attitudes into the profession. However, no psychometrically sound measure currently exists to assess students' perceptions of their attitudes toward the value of diversity in computing contexts nor their intentions to enact inclusive behaviors—both important indicators of the culture. Without such a measure, the research community lacks necessary tools for investigations into the effects that structural and curricular modifications have on building a more diverse and equitable computing culture. The purposes of the studies presented here are to detail validity, reliability and sensitivity assessments of a modified scale adapted from engineering, the VDEI-CS, designed to measure computer science undergraduate students' attitudes toward the value of diversity and intentions to enact inclusive behaviors.

This investigation includes two primary studies: Study 1 - the validation of the VDEI-CS, and Study 2 - a sensitivity study to determine if the VDEI-CS is indeed sensitive enough to detect changes in students' value of diversity and willingness to enact inclusive behaviors in computing contexts.

The following research questions (RQ) were addressed for Study 1 and Study 2, respectively:

- RQ 1: Does the VDEI-CS scale accurately measure students' attitudes toward the value of diversity and intentions to enact inclusive behaviors?

- RQ 2: Are the scores obtained using VDEI-CS sensitive enough to detect differences between students who participated in diversity promoting activities and those who did not?

Study 1 – Validation Study

Methods

Participants

A total of 149 first-year computer science majors enrolled at a R1 university completed the VDEI-CS at the beginning of their first year through an online platform. The sample of students represents a combined sample from three distinct cohorts – Cohort 1 beginning in 2017 (41%), Cohort 2 beginning in 2018 (34%) and Cohort 3 beginning in 2019 (26%). Furthermore, 34% of the total sample self-identified as women and 9% self-identified as being of an ethnicity typically underrepresented in STEM (URM). Participants in the study were enrolled in one of seven different computer science courses, spanning 12 classrooms and delivered by one of eight different instructors. The study was reviewed and approved by the Institutional Review Board (IRB# 1905584259) and participant consent was obtained through the online survey platform.

Data Analytic Approach

The VDEI-CS measure (see Appendix A) was modeled after the Valuing Diversity and Enacting Inclusion in Engineering (VDEIE) instrument geared towards engineering majors created by Rambo-Hernandez and colleagues [6]. The original VDEIE instrument was created due to the lack of psychometrically sound measures to assess engineering students' inclusive professional engineering identities through their attitudes toward the value of diversity and their intentions to enact inclusive behaviors in the context of engineering [6]. Nine items on the instrument remained unchanged. Minor adjustments to the wording of eight of the original seventeen items on the instrument were made such as replacing the word "engineering" with the phrase "computer science and game development." Two of these eight items on the original scale underwent slightly greater modifications but still involved primarily changing the word "engineering" to some form of "computer science and game development." For example, the original item, "*Engineers* should value diversity to increase public access to technology and *engineered* products" was replaced with "*Computer scientists and game developers* should value diversity to increase public access to technology, *computer science and game development* products."

The original VDEIE was rated on a 7-point Likert scale (1 = completely disagree to 7 = completely agree) and validity evidence was obtained through both EFA and CFA approaches [6]. The four-factor solution representing students' valuing of diversity to a) Serve Customers Better, and b) Fulfill a Greater Purpose and their willingness to act inclusively by a) Promoting Healthy Behaviors, and b) Challenging Discriminatory Behaviors on teams showed strong validity and reliability evidence both initially and across time ([6]).

Due to the valid and reliable nature of the VDEIE and its ability to accurately reflect students' standing on their value of diversity and willingness to act inclusively, the instrument was revised for computer science majors. Because the factor structure of the VDEIE had already been established, the minimally revised VDEI-CS instrument was subjected to CFA and reliability studies, hypothesizing a four-factor solution rated on a 7-point Likert scale, in conjunction with Rambo-Hernandez and colleagues' [6] previous findings. Given the clearly defined structure of the instrument and having an average of approximately nine participants per item, the sample size was adequate for validity investigations [31]-[33]

CFA was used as the primary analytical tool for the validity investigation. The size and significance of factor loadings onto their hypothesized factor was investigated at the standard $a = .05$ significance level. The overall model fit was evaluated using Hu and Bentler's [34] recommended global-fit statistics and cutoff values for an "adequate" fitting model (Root Mean Square Error Approximation $(RMSEA) < .08$, Comparative Fit Index $(CFI) > .9$, $Standardized\ Root\ Mean\ Square\ Residual\ (SRMR) < .08$). Chi square indices were also examined but can be subject to sample size and are not always a good indicator of model-fit [34].

Furthermore, noting the pooling of the data from three distinct cohorts for the validity investigation, it was important to investigate the invariance of the model across the cohorts. Noting the sample size, the Multiple Indicators Multiple Causes (MIMIC) structural equation modeling (SEM) approach—a form of CFA with covariates—was utilized for examining the invariance of indicators and latent means across the three cohorts [35], [36]. MIMIC models have been shown to be more appropriate for small samples (even if N = 150) than multiple-group CFA [37]. The invariance of the latent means was first tested by regressing the four latent factors on a cohort covariate. Insignificant regression coefficients would indicate invariance of the latent means across cohorts [36]. Next, differential item functioning within the MIMIC framework was utilized to examine the invariance of the indicator intercepts across cohorts. The four factors were regressed on the cohort covariate along with all of their indicators except one. The indicator for each factor used as the marker variable was not regressed onto the cohort covariate for model identification purposes [38]. These four indicator intercepts were tested for invariance in individual models. Insignificant effects of the cohort covariate on the indicators would indicate invariance of the item intercepts was upheld [36].

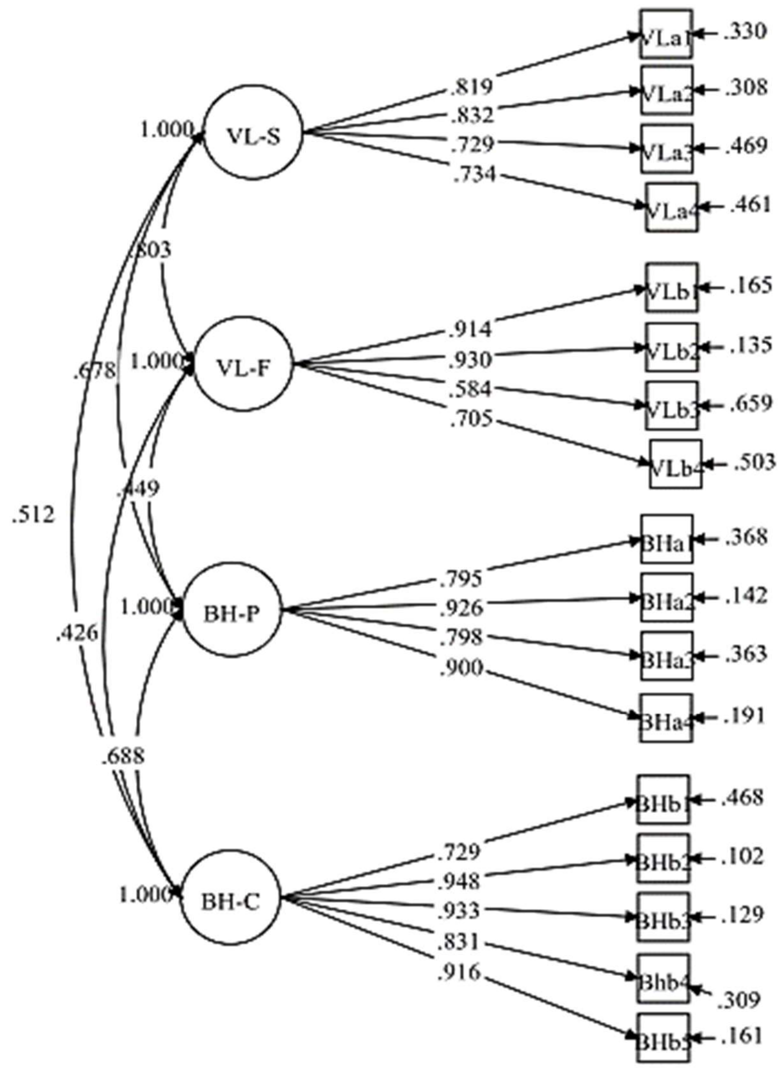
Lastly, the internal consistency of the instrument was evaluated for each factor using Cronbach's alpha. Mplus V8.7 [39] was used for CFA and SEM studies and STATA 17.1 [40] was used for descriptive, correlational and reliability studies.

Results

Descriptive statistics for each variable were observed and revealed some slight non normality to the data (see Appendix B). Thus, the CFA model and MIMIC models were estimated utilizing the Maximum Likelihood Robust estimation method (MLR) in Mplus, appropriate for data demonstrating slight non-normality [41]. A sample correlation matrix was also observed (see Appendix C).

CFA was used to confirm the internal structure of the VDEI-CS. The four-factor model demonstrated adequate fit ($X^2(113) = 205.09$, $p<.001$, RMSEA = .07, CFI = .92, and SRMR = .07) with all significant factor loadings for each item onto their hypothesized factor (see Figure 1). Correlations between all factors were also significant reflecting the multidimensional nature of the instrument (see Figure 1). Furthermore, all R^2 values were significant, with $p<.001$ for each of the four VDEI-CS factors, suggesting that for each observed variable a significant amount of its variance was explained by the underlying latent factor (ranged from .34 to .90). Modification indices suggested the cross loading of five items onto other factors with modification index values greater than 10 but less than 17. No changes were made to the instrument so as to maintain the distinguishable nature of the factors. These suggested cross loadings should be monitored in a larger, follow-up study.

Figure 1. Confirmatory Factor Analysis Results of VDEI-CS with Standardized Factor Loadings (all paths significant)



Note. Four latent factors are represented by circles: VL-S=Value Diversity to Serve Customers Better, VL-F=Value Diversity to Fulfill a Greater Purpose, BH-P=Enacting Inclusive Behaviors by Promoting Healthy Behaviors, and BH-C=Enacting Inclusive Behaviors by Challenging Discriminatory Behaviors. VLa1-BHb5 are measured variables for each factor that can be found in Appendix A.

Next, the MIMIC framework was used to investigate the invariance of the latent means across the three different cohorts by first regressing the four latent factors on a dummy-coded cohort covariate (1 = 2017 Cohort, 2 = 2018 Cohort, 3 = 2019 Cohort). Model results indicated an adequate fitting model ($X^2(126) = 229.06$, p<.001, RMSEA = .07, CFI = .92, and SRMR =

.07). The regression coefficients were all insignificant with p-values ranging from .17 to .73 (see Appendix D for the figure). These results indicated invariance of the latent means was upheld across the three cohorts [36].

Continuing the invariance investigation, differential item functioning within the MIMIC framework was utilized to examine the invariance of the indicator intercepts across cohorts. The four factors were regressed on the cohort covariate along with all but one of their indicators. Model results for the primary model indicated an adequate fit ($X^2(113) = 205.39$, $p<.001$, RMSEA = .07, CFI = .93, and SRMR = .07). Sixteen of the seventeen indicators yielded insignificant regression coefficients onto the cohort covariate with p-values ranging from .17 to .89 (for the primary model, see Appendix E). These results suggested that the invariance of the item intercepts was upheld across cohorts [36]. Item VLb3, however, demonstrated a significant p-value ($p = .017$) corresponding to its regression coefficient of .154. Thus, invariance of this item intercept was not upheld across cohorts and should be monitored in future investigations.

Cronbach's alpha was used to assess the internal consistency of the instrument. Each factor yielded good measures of internal consistency [32] as follows: Serve Customers Better (VL-S) - .86, Fulfill a Greater Purpose (VL-F) - .81, Promoting Healthy Behaviors (BH-P) - .91, and Challenging Discriminatory Behaviors (BH-C) - .93. The VDEI-CS, thus, appeared to accurately measure students' attitudes toward the value of diversity and intentions to enact inclusive behaviors in computing contexts.

Study 2 – Intervention Study

Methods

Participants

To assess if the VDEI-CS was sensitive enough to detect differences between students who participated in diversity promoting activities and those who did not, first-year students from Study 1 were divided into intervention and control groups. Though all 149 participants from Study 1 were originally included in Study 2, only 113 students completed the post-survey administered at the end of their first semester through the online platform. Retained students were clustered within 10 different classrooms with seven different instructors. Five classrooms were assigned to control status and the other five to the intervention status. Students assigned to intervention classrooms progressed through a series of intervention experiences over the course of the semester designed to enhance their attitudes regarding the valuing of diversity and inclusion within computing contexts. Such activities included: a Dean's welcome address that promoted inclusive behaviors within computing, reflective writing assignments, and an interactive theatre sketch. Student demographics for Study 2 were as follows: 85% intervention, 35% women and 9% URM.

Data Analytic Approach

Four separate two-level, random intercept hierarchical linear models (one for each scale) were constructed due to the nested nature of the data—students within classrooms. The intra-class correlations for each scale ranged from 0.00 to 0.23, yielding as little as 0% or as much as 23% of the variance for a particular scale was attributable to students' classrooms. A student's mean posttest score (level 1) was predicted by their pretest score and gender. The adjusted mean score of a classroom (level 2) was further predicted by the assigned instructor and intervention status of the classroom. Cross-level interactions were not modeled due to sample size limitations.

The level 1 model was depicted as:

$$y_{ij} = \beta_{0j} + \beta_{1j}Pretest_{ij} + \beta_{2j}Woman_{ij} + e_{ij},$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20} \quad (1)$$

Pretest represents student mean score on the related scale of the VDEI-CS at the beginning of the semester. *Woman* is a dummy-coded variable (1=woman, 0=man).

The level 2 model was given by:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}InstructorA_j + \gamma_{02}InstructorB_j + \gamma_{03}InstructorC_j + \gamma_{04}InstructorD_j + \gamma_{05}InstructorE_j + \gamma_{06}InstructorF_j + \gamma_{07}Intervention_j + \mu_{0j}. \quad (2)$$

InstructorA represents the instructor for the related course and is a dummy-coded variable (1=InstructorA, 0=other instructor). *InstructorB-InstructorF* are defined similarly. *Intervention* represents students' classroom intervention status and is a dummy-coded variable (1=intervention, 0=control).
The combined model was given by:

$$y_{ij} = \gamma_{00} + \gamma_{01}InstructorA_j + \gamma_{02}InstructorB_j + \gamma_{03}InstructorC_j + \gamma_{04}InstructorD_j + \gamma_{05}InstructorE_j + \gamma_{06}InstructorF_j + \gamma_{07}Intervention_j + \gamma_{10}Pretest_{ij} + \gamma_{20}Woman_{ij} + e_{ij} + \mu_{0j}\,. \quad (3)$$

Of particular importance, γ_{07} represents the effect of intervention status on mean classroom postscores while controlling for other variables.
STATA 17.1 [40] was used for descriptive statistics and model estimations.

Results

Descriptive statistics of the four VDEI-CS scales for pre and posttest administrations are provided in Table 1.

Table 1. Pre and post VDEI-CS descriptive statistics for intervention and control groups

	Time	Intervention			Control		
		N	**Mean**	**SD**	***N***	**Mean**	**SD**
Serve Customers Better	Pretest	96	6.19	.94	16	5.64	1.12
	Posttest	96	6.13	.95	17	5.23	1.69
Fulfill Greater Purpose	Pretest	96	5.91	1.09	16	5.42	1.39
	Posttest	96	5.96	1.02	17	5.12	1.70
	Pretest	96	6.24	.90	17	6.22	.79

Promoting Healthy Behaviors	Posttest	96	6.31	.87	17	6.14	1.38
Challenging Discriminatory Behaviors	Pretest	94	6.05	1.18	17	6.01	1.04
	Posttest	96	6.00	1.28	17	5.71	1.91

To determine if the intervention and control groups were indistinguishable on each VDEI-CS scale pretest scores, Mann-Whitney U Tests utilizing a Bonferroni correction were employed. All related p-values were greater than the Bonferroni corrected significance value of .0125. Though average scores on the VDEI-CS pretest scales were indistinguishable between intervention and control groups, it was determined to utilize student pretest scores as a control variable in the hierarchical linear modeling technique to allow for the most accurate estimations of the effect of interventions on student post-scores.

The validity of the results of intervention investigation are dependent upon the satisfaction of the assumptions of the analytical modeling technique. In review of these assumptions, slight violations were discovered particularly regarding non-normality of level 1 residuals and heterogeneity of level 1 residual variances. These violations were addressed by utilizing a robust estimator for standard errors [40].

Parameter estimates for the four models representing the four VDEI-CS scales are provided in Table 2.

Table 2. Parameter estimates for the four VDEI-CS models

	Serve Customers Better	**Fulfill a Greater Purpose**	**Promoting Healthy Behaviors**	**Challenging Discriminatory Behaviors**
Fixed Effects				
Intercept (γ_{00})	2.02 (.85)*	1.66 (.83)*	2.00 (.67)**	2.28 (.68)**
InstructorA (γ_{01})	.70 (.15)***	.41 (.03)***	-.04 (.14)	-.39 (.06)***
InstructorB (γ_{02})	.46 (.03)***	.13 (.03)***	-.19 (.11)	-.42 (.08)***
InstructorC (γ_{03})	.17 (.09)	-.26 (.19)	-.94 (.16)***	-.14 (.32)
InstructorD (γ_{04})	.19 (.79)	.00 (.35)	.92 (.22)***	-1.20 (.40)**
InstructorE (γ_{05})	.95 (.11)***	.49 (.07)***	.14 (.18)	-.68 (.21)**
InstructorF (γ_{06})	1.65 (.80)*	1.20 (.42)**	.70 (.21)**	-.41 (.32)
Intervention (γ_{07})	.95 (.81)	.71 (.34)*	.79 (.14)***	-.07 (.13)
Pretest (γ_{10})	.41 (.13)**	.55 (.17)**	.59 (.12)***	.70 (.14)***
Woman (γ_{20})	-.03 (.13)	.06 (.11)	-.17 (.19)	-.04 (.27)
Random Effects				
$Residual\ (\sigma_e^2)$	.93 (.32)	.85 (.28)	.59 (.16)	1.05 (.32)
Classroom (σ_μ^2)	.01 (.02)	<.001	<.001	<.001

Note. ***p<.001, **p<.01, *p<.05

The investigation of *RQ2* revealed a statistically significant difference between intervention and control groups on both the Fulfill a Greater Purpose and Promoting Healthy Behaviors scales. On average, the Fulfill a Greater Purpose post-scale scores were .71 points higher for classrooms involved in intervention experiences compared to classrooms who were not involved in intervention experiences, controlling for other variables. Similarly, the Promoting Healthy Behaviors post-scale scores were, on average, .79 points higher for classrooms who participated in intervention experiences compared to classrooms who did not participate in intervention experiences, controlling for other variables. These effects were after accounting for the effect of instructor. Also, worthy of noting, after accounting for intervention and instructor, there was essentially no variability between classrooms left to be explained across the means of the four scales. It appeared that the VDEI-CS was sensitive enough to detect changes in student attitudes towards the value of diversity and willingness to act inclusively over the course of the semester, but there was no evidence that the instrument was sensitive enough to detect changes in their intentions to serve customers better or challenge discriminatory behavior.

Discussion

This two-part study revealed several findings of interest. The first study which included investigations into the factor structure and internal consistency of the VDEI-CS yielded good validity and reliability evidence for the four-factor instrument. The four-factor structure of the instrument was validated through CFA and MIMIC modeling to ensure invariance across the sampling cohorts. This finding further substantiates the validity of the original VDEIE instrument created by Rambo-Hernandez and colleagues [6] for engineering students and extends its applicability to computer science students. The computing research community now has an instrument that can accurately measure computing students' value of diversity and willingness to enact inclusive behaviors within computing contexts. This is a pertinent tool for diversity investigations in computer science.

Focusing on the results of the second study, prior to any direct interventions, the results indicated that intervention and control groups did not differ at pretest across any of the factors. This lack of a difference prior to interventions is a positive indication of the validity of the survey, as one would not expect students in the intervention and control classrooms to respond differently before participating in any interventions. As the semester progressed, intervention classrooms demonstrated statistically significantly higher scores on the Willingness to Act Inclusively by Promoting Healthy Behaviors scale of the VDEI-CS, which mimics findings by Rambo-Hernandez and colleagues [6]. In a longitudinal investigation of engineering student VDEIE scale scores, Rambo-Hernandez and colleagues [6] discovered that students in the intervention group demonstrated statistically significantly greater increases in their Willingness to Act Inclusively by Promoting a Healthy Work Environment (later deemed "Promoting Healthy Behaviors"). In both studies, intervention practices appear to influence some portion of students' willingness to act inclusively, though differences in the current study were also found in students' valuing of diversity to fulfill a greater purpose due to intervention experiences. Though the VDEI-CS detected changes in these two factors between intervention and control classrooms, there was no evidence that the instrument was sensitive enough to detect changes in students' intentions to serve customers better or challenge discriminatory behaviors. More study is needed.

Implications for Future Research

The establishment of a psychometrically sound instrument to measure students' valuing of diversity and willingness to act inclusively in computer science settings holds great importance for future investigations. Noting the extreme lack of diversity in computer science degree programs and fields, it is important that energy be given towards cultivating an inclusive and equitable culture. Proper use of the VDEI-CS will allow researchers and practitioners to begin investigating computer science students' attitudes towards the value of diversity and their willingness to act inclusively in computing contexts, developmental patterns to these attitudes and actions, and factors (including interventions) that influence (positively or negatively) the cultivation of positive attitudes and actions towards diversity in computer science.

Limitations

There are several limitations to this study that need to be addressed. First, the sample size for Study 1, though adequate was not optimal and spanned three cohorts. CFA modification indices suggested the cross-loading of a few items. MIMIC modeling also suggested DIF existed for one item. These hindrances can each be attributed to the sample size and technique. The VDEI-CS needs to be further validated with a larger and more diverse sample to determine if these findings need to be further addressed.

The sample size for Study 2 was also small with students only nested within ten classrooms, some of which held as few as two students. The design effects were also less than 1.10 for two of the four factors whose intraclass correlations were 0.00, suggesting that a single-level analysis might be sufficient and not lead to overly misleading results [42]. However, as best practice indicates, we proceeded with a multilevel modeling analytical approach for all scales to account for any variability in scores due to clustering and maintain consistency of the analysis across the scales. Furthermore, the lack of meeting all of the assumptions for multilevel modeling was addressed by using the robust standard estimator. Though sufficient, this is not optimal. A larger study is needed to support the findings from Study 2.

For this study, instructors either taught only intervention courses, or only non-intervention courses. Though attempts were made to control for instructor effects, the confounding of this variable with the intervention variable makes it difficult to completely determine the effect of the intervention experiences alone on postscores between intervention and control classrooms. However, this study is a preliminary investigation. A larger study with a more appropriate experimental design is needed to validate the effects of the intervention experiences on students' valuing of diversity or willingness to act inclusively.

Conclusion

The computing culture has long been characterized by a lack of diversity. Noting the high and increasing demand for computer scientists combined with the benefits of a diverse computing workforce that broadens the computing research agenda, fosters creativity, provides new perspectives on problems, and facilitates the construction of new, equity-centered, technologies, working to create a more diverse computing culture is a necessary endeavor. The absence of voices of students and computing professionals from non-majority groups drastically limits the positive impact computer scientists can have and the scope of new technologies they can develop. The VDEI-CS scale could be used to assess student attitudes toward the value of diversity and their intentions to enact inclusive behaviors. Having a psychometrically sound tool to assess student attitudes toward the value of diversity and inclusive behaviors in computing contexts will enable researchers to gauge the temperature of a group of students and assess the effect of interventions developed to promote change within the culture.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant # 2033129. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] U.S. Bureau of Labor Statistics, "Computer and information technology occupations: Occupational outlook handbook: U.S. bureau of labor statistics 2022," *Bls.gov*, https://www.bls.gov/ooh/computer-and-information-technology/home.htm.

[2] National Science Foundation, "Women, minorities, and persons with disabilities in science and engineering: 2021," *ncses.nsf.gov*. https://ncses.nsf.gov/pubs/nsf21321/report/about-this-report.

[3] Department for Education, "*Revised A level and other 16-18 results in England*," 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/676389/SFR03_2018_Main_text.pdf.

[4] T. J. Bright et al., "Making the case for workforce diversity in biomedical informatics to help achieve equity-centered care: A look at the AMIA First Look Program," *Journal of the American Medical Informatics Association*, vol. 29, no. 1, pp. 171–175, Nov. 2021, doi: 10.1093/jamia/ocab246.

[5] S. E. Page, *The difference: how the power of diversity creates better groups, firms, schools, and societies.* Princeton University Press, Princeton, NJ, 2007.

[6] K. E. Rambo-Hernandez et al., "Valuing diversity and enacting inclusion in engineering (VDEIE): Validity evidence for a new scale," *International Journal of Engineering Education,* vol. 37, no. 5, pp. 1382-1397, May 2021.

[7] I. Ajzen and M. Fishbein, Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ, 1980.

[8] S. Beyer. "Predictors of female and male Computer Science students' grades." *Journal of Women and Minorities in Science and Engineering,* vol. *14*, pp. 377–409, 2008. .10.1615/JWomenMinorScienEng.v14.i4.

[9] S. Beyer, "Gender differences and intra-gender differences amongst Management Information Systems students," *Journal of Information Science Education,* vol. *19*, pp. 301–310, 2008.

[10] S. Beyer, "Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades," *Computer Science Education*, vol. 24, no.2–3, pp. 153–192, Aug. 2014. https://doi-org.srv-proxy1.library.tamu.edu/10.1080/08993408.2014.963363.

[11] S. M. Haller and S. Beyer, "Gender differences and intragender differences in computer science students: Are female CS majors more similar to male CS majors or female nonmajors," *Journal of Women and Minorities in Science and Engineering*, vol. 12, pp. 337-365, 2006.

[12] J. M. Cohoon and W. Aspray, *Women and information technology: Research on underrepresentation*. Cambridge, Mass.; London: Mit, 2008.

[13] S. J. Correll, "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations," *American Sociological Review*, vol. 69, no. 1, pp. 93–113, Feb. 2004, doi: 10.1177/000312240406900106.

[14] J. L. Cundiff, T. K. Vescio, E. Loken, and L. Lo, "Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors?," *Social Psychology of Education*, vol. 16, no. 4, pp. 541–554, Jun. 2013, doi: 10.1007/s11218-013-9232-8.

[15] C. Good, J. Aronson, and J. A. Harder, "Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses," *Journal of Applied Developmental Psychology*, vol. 29, no. 1, pp. 17–28, Jan. 2008, doi: 10.1016/j.appdev.2007.10.004.

[16] C. M. Jagacinski, "Women Engineering Students: Competence Perceptions and Achievement Goals in the Freshman Engineering Course," *Sex Roles*, vol. 69, no. 11–12, pp. 644–657, Sep. 2013, doi: 10.1007/s11199-013-0325-9.

[17] S. K. Maheshwari, A. L. Pierce, L. Anne and E. G. Zapatero, "Understanding the lack of minority representation in graduate programs in computer science and information technology: A focus group study of student perceptions," *Academy of Information & Management Sciences Journal*, vol. 12, no. 1-2, pp. 71-90, 2009.

[18] J. G. Stout, N. Dasgupta, M. Hunsinger, and M. A. McManus, "STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM)," *Journal of personality and social psychology*, vol. 100, no. 2, pp. 255–270, 2011. https://doi.org/10.1037/a0021385

[19] A. Szymanowicz and A. Furnham, "Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta-analyses," *Learning and Individual Differences*, vol. 21, pp. 493-504, 2011.

[20] G. M. Walton, G. L. Cohen, D. Cwir, and S. J. Spencer, "Mere belonging: The power of social connections," *Journal of Personality and Social Psychology*, vol. 102, no. 3, pp. 513–532, 2012, doi: 10.1037/a0025731.

[21] G. M. Walton and G. L. Cohen, "A question of belonging: Race, social fit, and achievement," *Journal of Personality and Social Psychology*, vol. 92, no. 1, pp. 82–96, 2007, doi: 10.1037/0022-3514.92.1.82.

[22] C. Convertino and A. Monarrez, "Storylines in figured worlds: Understanding diverse college students' decision to major in computer science," *Journal of Diversity in Higher Education*, Feb. 2020, doi: 10.1037/dhe0000172.

[23] J. M. Cohoon, "Recruiting and retaining women in undergraduate computing majors," *ACM SIGCSE Bulletin*, vol. 34, no. 2, pp. 48–52, Jun. 2002, doi: 10.1145/543812.543829.

[24] Y. Ben-David Kolikant, “Computer science education as a cultural encounter: A socio-cultural framework for articulating teaching difficulties,” *Instructional Science*, vol. 39, no. 4, pp. 543–559, Jun. 2010, doi: 10.1007/s11251-010-9140-7.

[25] M. Ong, N. Jaumot-Pascual, and L. T. Ko, “Research literature on women of color in undergraduate engineering education: A systematic thematic synthesis,” *Journal of Engineering Education*, vol. 109, no. 3, pp. 581–615, Jul. 2020, doi: 10.1002/jee.20345.

[26] C. Seron, S. S. Silbey, E. Cech, and B. Rubineau, “Persistence Is Cultural,” *Work and Occupations*, vol. 43, no. 2, pp. 178–214, Dec. 2015, doi: 10.1177/0730888415618728.

[27] E. Seymour and N. M. Hewitt, *Talking about leaving: Why undergraduates leave the sciences*. Boulder (Colo.): Westview Press, 2000.

[28] A. Godwin and G. Potvin, “Pushing and pulling Sara: A case study of the contrasting influences of high school and university experiences on engineering agency, identity, and participation,” *Journal of Research in Science Teaching*, vol. 54, no. 4, pp. 439–462, Nov. 2016, doi: 10.1002/tea.21372.

[29] J. M. Cohoon, “Toward improving female retention in the computer science major,” *Communications of the ACM*, vol. 44, no. 5, pp. 108–114, May 2001, doi: 10.1145/374308.374367.

[30] M. C. Murphy, C. M. Steele, and J. J. Gross, “Signaling Threat,” *Psychological Science*, vol. 18, no. 10, pp. 879–885, Oct. 2007, doi: 10.1111/j.1467-9280.2007.01995.x.

[31] P. M. Bentler and C.-P. Chou, “Practical Issues in Structural Modeling,” *Sociological Methods & Research*, vol. 16, no. 1, pp. 78–117, Aug. 1987, doi: 10.1177/0049124187016001004.

[32] P. Kline, *The handbook of psychological testing*. London; New York: Routledge, 1999.

[33] L. Hatcher, *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, Nc: Sas Institute, 1994.

[34] L. Hu and P. M. Bentler, “Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives,” *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 6, no. 1, pp. 1–55, Jan. 1999, doi: 10.1080/10705519909540118.

[35] L. K. Muthén and B. O. Muthén, *Mplus Short Courses, Topic 1: Exploratory Factor Analysis, Confirmatory Factor Analysis, and Structural Equation Modeling for Continuous Outcomes.* Los Angeles, CA: Muthén & Muthén, 2009.

[36] J. Wang and X. Wang, *Structural equation modeling: Applications using Mplus*. Hoboken, Nj: Wiley, 2020.

[37] T. A. Brown, *Confirmatory factor analysis for applied research*. New York: Guilford, 2015.

[38] D. Kaplan, *Structural equation modeling: Foundations and extensions*. Los Angeles: Sage, 2009.

[39] L. K. Muthén and B. O. Muthén, *Mplus user's guide (eighth edition).* Muthén & Muthén. 1998-2021.

[40] StataCorp, *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC. 2021.

[41] G. R. Hancock and R. O. Mueller, *Structural equation modeling: A second course*. Charlotte, Nc: Information Age Publishing, Inc, 2013.

[42] M. H. C. Lai and O. Kwok, "Examining the rule of thumb of not using multilevel modeling: The "Design effect smaller than two" rule," *The Journal of Experimental Education*, vol. 83, no. 3, pp. 423-438, doi: 10.1080/00220973.2014.907229

Appendix A

Valuing Diversity and Enacting Inclusion in Computer Science

Computer scientists and game developers should value diversity to:

Serve Customers Better (VL-S factor with VL-a items 1-4 below)

1. Help them understand client and customer needs.
2. Improve products.
3. Increase public access to technology, computer science and game development products.
4. Collaborate effectively with stakeholders in a computer science and game development project.

Fulfill a Greater Purpose (VL-F factor with VL-b items 1-4 below)

1. Fulfill a social responsibility for making the world better.
2. Work for a greater cause.
3. Help improve the bottom line.
4. Do the right thing.

While working on a team, I:

Promote Healthy Behaviors (BH-P factor with BH-a items 1-4 below)

1. Include everyone in all team meetings.
2. Make sure to give credit to team members who make contributions to the project.
3. Make sure all team members have the opportunity to take part in decision-making.
4. Make sure every team member has the opportunity to contribute to the project.

Challenge Discriminatory Behaviors (BH-C factor with BH-b items 1-5 below)

1. Challenge homophobic behaviors.
2. Challenge racist behaviors.
3. Challenge any type of discriminatory behaviors.
4. Challenge sexist behaviors.
5. Challenge xenophobic behaviors, which are behaviors that discriminate against people from other countries.

Appendix B

Descriptive Statistics for the 17 VDEI-CS Items Rated on a 7-Point Likert Scale (n=149)

Variable	Mean	SD	Skewness	Kurtosis
VLa1	5.99	1.23	-1.61	6.13
VLa2	6.11	1.15	-1.70	6.38
VLa3	6.22	1.05	-1.75	6.24
VLa4	6.01	1.18	-1.28	4.34
VLb1	5.83	1.32	-1.30	4.75
VLb2	5.97	1.24	-1.56	5.98
VLb3	5.71	1.39	-1.25	4.52
VLb4	5.74	1.41	-1.19	3.84
BHa1	6.05	1.05	-1.38	4.90
BHa2	6.21	.94	-1.11	3.57
BHa3	6.48	.79	-1.58	5.02
BHa4	6.24	.93	-1.54	5.95
BHb1	5.81	1.48	-1.59	2.28
BHb2	6.20	1.17	-1.97	7.51
BHb3	6.10	1.21	-1.65	6.06
BHb4	6.17	1.17	-1.49	4.62
BHb5	6.28	1.09	-1.81	6.55

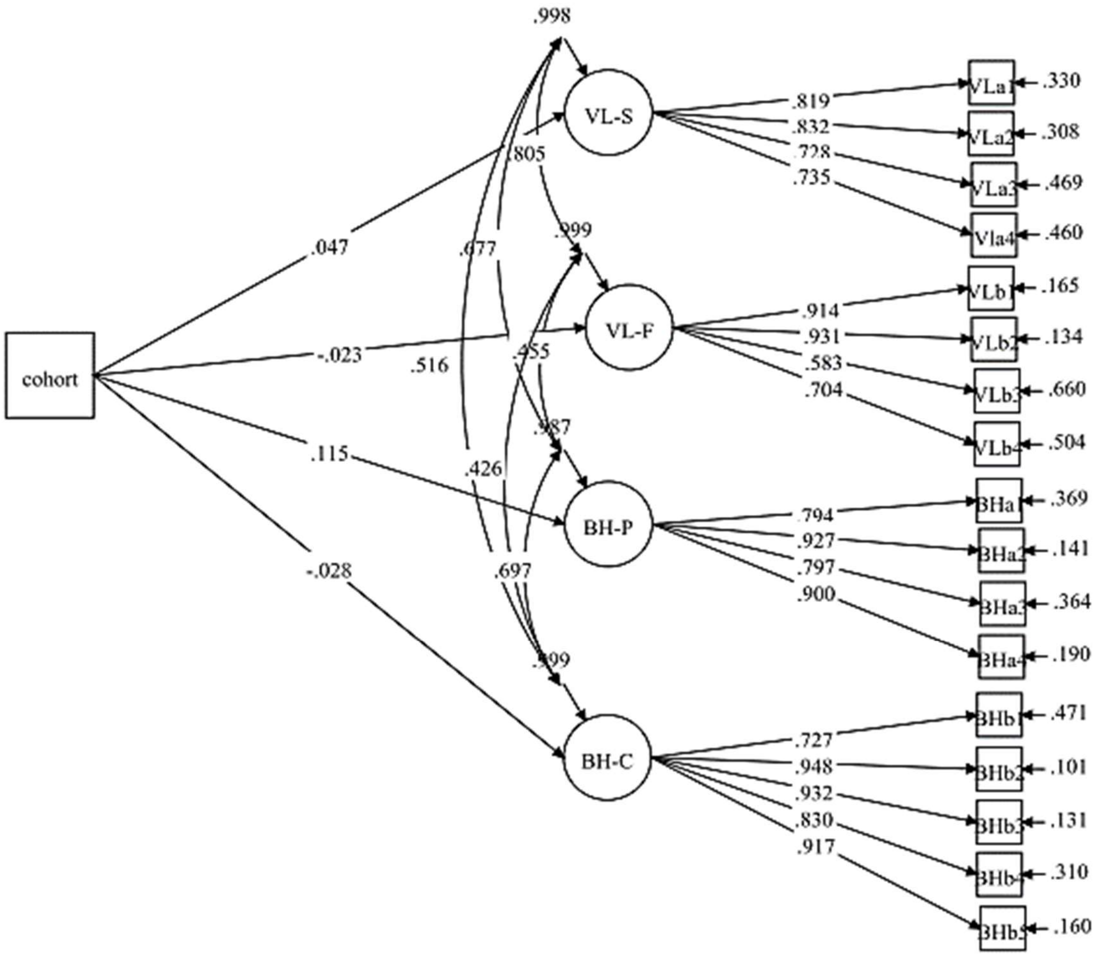
Appendix C

Correlation Matrix for the 17 VDEI-CS Items

	VLa1	VLa2	VLa3	VLa4	VLb1	VLb2	VLb3	VLb4	BHa1	BHa2	BHa3	BHa4	BHb1	BHb2	BHb3	BHb4	BHb5
VLa1	1.00																
VLa2	.75	1.00															
VLa3	.64	.54	1.00														
VLa4	.71	.61	.86	1.00													
VLb1	.45	.50	.48	.51	1.00												
VLb2	.44	.38	.67	.62	.47	1.00											
VLb3	.50	.55	.48	.51	.61	.56	1.00										
VLb4	.49	.61	.40	.50	.61	.46	.72	1.00									
BHa1	.33	.32	.32	.25	.29	.28	.38	.30	1.00								
BHa2	.42	.39	.29	.26	.40	.33	.43	.36	.42	1.00							
BHa3	.46	.49	.35	.34	.42	.34	.52	.48	.49	.74	1.00						
BHa4	.32	.36	.39	.28	.36	.38	.44	.41	.67	.52	.57	1.00					
BHb1	.33	.36	.36	.27	.33	.39	.42	.40	.63	.56	.61	.90	1.00				
BHb2	.34	.36	.36	.27	.33	.39	.42	.40	63	.50	.60	.77	.79	1.00			
BHb3	.46	.43	.33	.28	.34	.35	.48	.39	.37	.64	.74	.49	.48	.52	1.00		
BHb4	.44	.50	.38	.39	.41	.39	.56	.53	.41	.71	.83	.52	.58	.58	.73	1.00	
BHb5	.45	.48	.48	.38	.39	.42	.52	.45	.70	.60	.69	.87	.82	.76	.62	.66	1.00

Appendix D

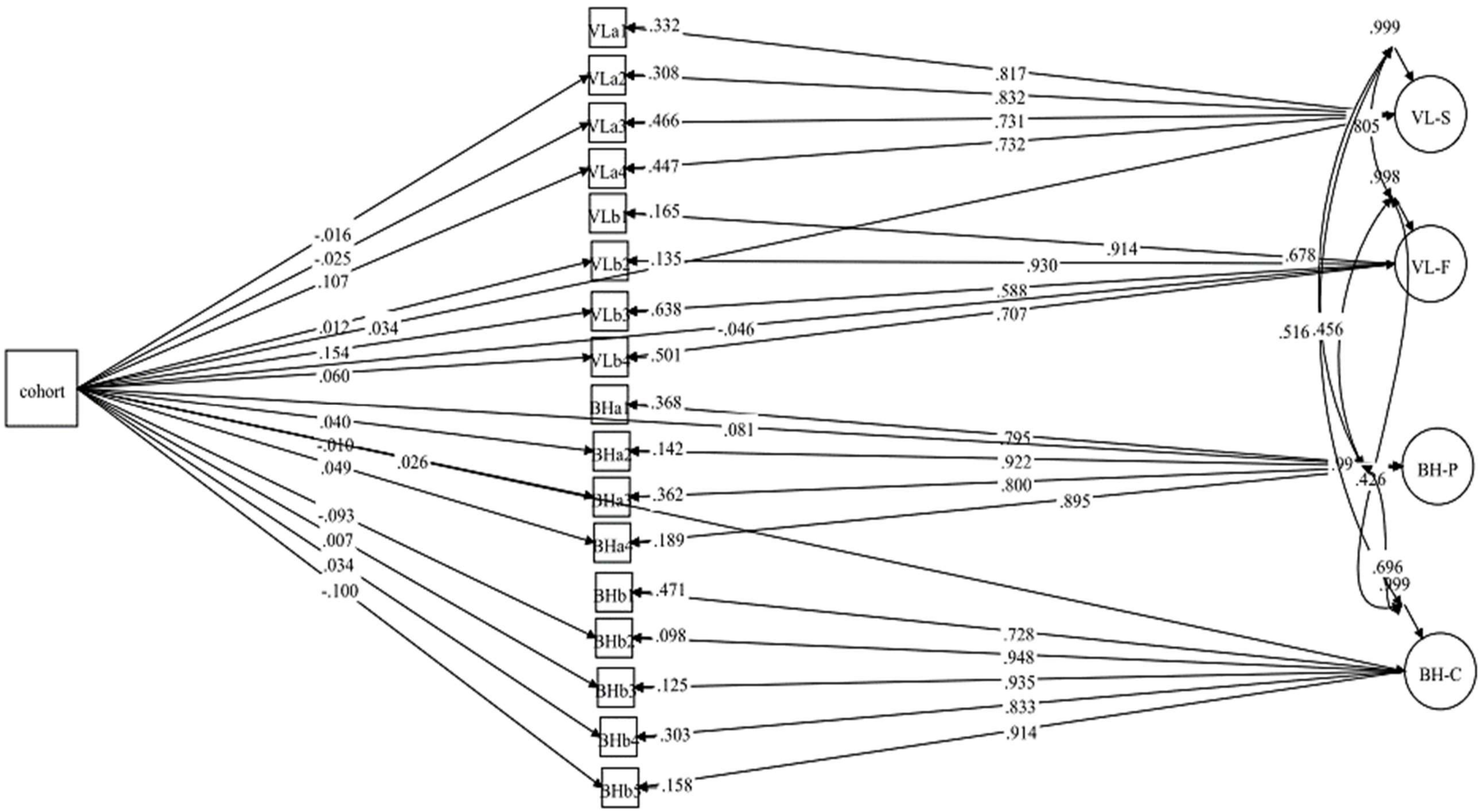
MIMIC Model of the Four-Factor VDEI-CS across Cohorts



Note. Four latent factors are represented by circles: VL-S=Value Diversity to Serve Customers Better, VL-F=Value Diversity to Fulfill a Greater Purpose, BH-P=Enacting Inclusive Behaviors by Promoting Healthy Behaviors, and BH-C=Enacting Inclusive Behaviors by Challenging Discriminatory Behaviors. VLa1-BHb5 are measured variables for each factor that can be found in Appendix A.

Appendix E

Testing Differential Item Functioning using MIMIC Modeling for the DVEI-CS



Note. Four latent factors are represented by circles: VL-S=Value Diversity to Serve Customers Better, VL-F=Value Diversity to Fulfill a Greater Purpose, BH-P=Enacting Inclusive Behaviors by Promoting Healthy Behaviors, and BH-C=Enacting Inclusive Behaviors by Challenging Discriminatory Behaviors. VLa1-BHb5 are measured variables for each factor that can be found in Appendix A