Abstract

Kolb's experiential learning cycle theorizes the process of learning through a hands-on experience. Although the Association of American Colleges and Universities (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics provide a direct measure of the qualities of this learning cycle, few indirect measures have been developed to accompany the rubrics and the learning cycle. This paper aims to demonstrate construct validity and measurement invariance of pre-experience and post-experience surveys intended to measure undergraduate students' perception of learning in an experiential learning context. Construct validity and longitudinal measurement invariance were examined through a confirmatory factor analysis. Findings suggest the instruments provide an adequate measure of students' perceptions of learning. In addition, partial scalar measurement invariance was achieved supporting the ability to compare growth between surveys. The survey instruments serve as strong indirect measures of Experience Learning Student Learning Outcomes. These indirect measures, coupled with direct measures, provide evidence of learning through hands-on experiences, however evidence of growth is less robust.

AUTHORS

John P. Walker, MPS
University of Tennessee

Louis M. Rocconi, Ph.D.
University of Tennessee

Experiential Learning Student Surveys: Indirect Measures of Student Growth

A steady increase of scholarship in experiential learning has demonstrated a need for effective measures of these experiences (Seaman et al., 2017). Boyatzis et al. (1995) proposed that institutions conduct longitudinal studies to determine the value added to learning and continuously assess the learning process. Although there are several methods to collect student data to measure learning in an experiential context, indirect measures, such as surveys, allow students to reflect on the learning experience from their perspective (Banta & Palomba, 2015). Surveys used in a within-subjects design provide vital information on growth; however, researchers and practitioners need to demonstrate that these surveys measure what they intend to measure and do so consistently. This paper assesses the construct validity and measurement invariance of two surveys, pre-experience and post-experience, used to measure student attainment of the Experience Learning Student Learning Outcomes (SLOs), as established by a southeastern U.S., four-year, research university.

CORRESPONDENCE

Email
jwalk141@utk.edu

Kolb's Experiential Learning Cycle

The notion of learning by experience is not a new concept. Notable educational analysts, John Dewey and David Kolb, each laid the groundwork for the importance of experiential learning. Dewey (1938) contended that students' potential is hindered by the traditional classroom approach to learning which focused on delivering knowledge and little emphasis on the application of the knowledge. Kolb (1984) formalized the fundamental theory of experiential education which summarized the steps needed for learning to happen in a hands-on experience. Kolb posited, “learning is the process whereby knowledge is created through the transformation of experience” (Kolb, 1984, p. 38). He theorized that experience, coupled with structured reflection, allows students to grow through hands-on learning-a cyclical process that outlines the pathway to learning by doing, reviewing, concluding, and planning. Reflection and thoughtful planning follow each experience which leads to improvements for the next experience.

Kolb’s experiential learning model is based on his identification of two ways of *acquiring knowledge and skills* through an experience (concrete experience and abstract conceptualization) and two ways of *transforming* through an experience (reflective observation and active experimentation). The first step of Kolb’s experiential learning cycle states that learning begins by doing. Students will only begin to construct the skills and knowledge needed through observation and application in an experience. The first step serves as the cornerstone to the learning process. Once students have completed the initial experience, reflective behavior pushes learning forward. Reflection, the second step, allows students to strategize improvements and brainstorm new ideas for the next experience. The third step, abstract conceptualization, occurs when students formulate solutions for improvement to apply to the next experience based on their reflection. Students can then implement the solution as the final step of the learning process through experimentation. The iterative process continues through further reflection, planning, and testing that builds on one experience after the next.

Kolb’s experiential learning theory has been adapted and applied in several higher education contexts. The theory has been applied to many course-based curricula (Abdulwahed & Nagy, 2009; Healey & Jenkins, 2000; Petkus, 2000; Russell-Bowie, 2013) and aligned to other educational models in higher education (Baker et al., 2012; Poore et al., 2014). For instance, Reshmad’sa and Vijayakumari (2017) investigated the pedagogical skills of pre-service teachers (n = 40) in student teaching roles. The authors measured students’ teaching aptitude and their use of active learning strategies in a classroom setting. They found the use of Kolb’s experiential learning strategy was substantially more effective than conventional teaching strategies. That is, student teachers that used Kolb’s experiential learning strategies were more reflective and demonstrated better development of pedagogical skills compared to student teachers using conventional teaching strategies.

....experience, coupled with structured reflection, allows students to grow through hands-on learning

Some researchers have attempted to directly measure experiential learning opportunities based on Kolb’s experiential learning theory model. For example, Smith and Rayfield (2017) used Kolb’s Learning Style Inventory (KLSI) to examine the preferred learning styles of individuals in an experiential setting. The KLSI is a direct measure designed to categorize students into nine learning styles related to Kolb’s learning model. However, the inventory has received criticism for its “pigeonhole” approach to evaluate learner style (Manolis et al., 2013). The KLSI is focused on identifying an individual’s learning style with little emphasis on the measure of actual learning. Additionally, the inventory does not take learner perception of self-efficacy into account. An indirect measure of learner perceptions could provide a new perspective of self-assessment and self-awareness through the learning process.

Some studies have noted that Kolb’s experiential learning cycle falls short of capturing the learning process through hands-on experiences (Bergsteiner et al., 2010; Miettinen, 2010). Bergsteiner and Avery (2014) suggested that the model insufficiently attends to the numerous facets of real-world learning including cultural and emotional contexts that can be captured through student perceptions of their learning environment. Warren et al. (1995) argued, “experiential methodology is not linear, cyclical, or even patterned. It is a series of working principles, all of which are equally important or must be present to varying degrees at some time during experiential learning” (p. 243). Through an examination of the literature, Chapman et al. (1992) found experiential education is grounded on, among other principles, meaningful relationships built from collaborative endeavors, structured reflection, and emotional investment and engagement in the experience. Although Kolb and colleagues have examined the role of teamwork and collaboration in experiential learning in later work (Kayes et al., 2005), the cycle fails to emphasize the importance of collaboration, and the principles of lifelong learning and engaged scholarship. Thus, we sought to develop a valid and reliable instrument that incorporated these missing learning characteristics into the survey instruments.

AAC&U's VALUE Rubrics

Two experiential learning survey instruments, based on research and instruments from the Association of American Colleges and Universities (AAC&U), were developed to measure undergraduate students' perception of their learning in an experiential learning context. The AAC&U created 16 rubrics to measure student learning on skills essential to employers and faculty known as the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics (Rhodes, 2010). Several VALUE rubrics measure skills essential to experiential learning theory (e.g., critical thinking, foundations and skills for lifelong learning, teamwork, integrative learning) and have been widely used across many institutions to measure these latent traits objectively and reliably (Finley, 2011; Rhodes, 2010; Rhodes & Finley, 2013). They are designed to be a direct measure of student achievement, but modifying the language allows the rubrics to be used as an indirect measure to quantify learning from the learners' perspective. The rubrics served as an important framework to develop the Experiential Learning Student Surveys. The surveys include measures on teamwork (collaboration), lifelong learning, and engagement, elements not explicitly mentioned in Kolb's experiential learning cycle. Table 1 shows how each item in the surveys is aligned with the VALUE rubrics.

The purpose of this study was to assess the construct validity and longitudinal measurement invariance of two surveys designed to measure students' perceptions of learning in an experiential learning context.

Experience Learning SLOs

The primary goal of experiential learning is to enhance students' development and educational experiences by providing more opportunities for real-world learning. Experiential learning is most effective when it is a dynamic approach in which students engage, apply, collaborate, and reflect on course content and lessons learned (Kolb & Kolb, 2011). The Experience Learning SLOs were designed to incorporate Kolb's four-stage learning cycle with the addition of collaboration and lifelong learning. Since learning occurs at all of these stages, it is important to measure students' learning and growth throughout the process. These stages of experiential learning therefore formed the foundation for defining our desired SLOs:

SLO 1: Students will value the importance of engaged scholarship and lifelong learning.

SLO 2: Students will apply knowledge, values, and skills in solving real-world problems.

SLO 3: Students will work collaboratively with others.

SLO 4: Students will engage in structured reflection as part of the inquiry process.

The Experience Learning SLOs represent a holistic approach to learning that emphasizes learning through experiences. The four interrelated SLOs are assessed using two indirect measures (pre-experience and post-experience surveys, described later) to understand student perceptions of self-efficacy in an experiential learning context.

Purpose

The purpose of this study was to assess the construct validity and longitudinal measurement invariance of two surveys designed to measure students' perceptions of learning in an experiential learning context. Specifically, we were interested in measuring students' attainment of the above-mentioned Experience Learning SLOs. We utilized confirmatory factor analysis to assess the extent to which the pre-experience and post-experience surveys measure the SLOs.

Table 1.
Alignment of survey items with Experience Learning SLOs and AAC&U VALUE rubric

Item No.	Pre-Experience Survey Items	Post-Experience Survey Items	SLO	VALUE Rubric Item	VALUE Rubric
1	I often participate in activities that serve the needs of others	I am interested in exploring the problems of society (i.e. the needs of others)	1	Show evidence of interest in the problems of society (needs of others)	Foundations and Skills for Lifelong Learning
2	I think it is important for the university to use its resources for the benefit of society	I think it is important for academia to use their resources for the benefit of society	1	Value (i.e., offer a positive attitude toward) the use of engaged scholarship to address societal problems	Foundations and Skills for Lifelong Learning
3	I often participate in academic activities/events that aim to help others	I am interested in using the skills and knowledge that I have acquired from this course to contribute to the public good	1	Demonstrate a desire to utilize engaged scholarship	Civic Engagement
4	I typically like to explore more than usual when I am learning something new that interests me	I want to continue to develop relevant skills that are related to this experience	1	Demonstrate a commitment to lifelong learning	Foundations and Skills for Lifelong Learning
5	I can clearly describe a real world problem related to this course to someone that knows little about the problem	I can clearly describe a real-world problem related to this course to someone that knows little about the problem	2	Clearly describe a real-world problem amenable to engaged scholarship	Critical Thinking
6	I have been introduced to more than one way to address real-world problem(s) related to this course	I have been introduced to more than one way to address real world problem(s) that my faculty member/professor brought up in this course.	2	Analyze literature (content/research methods) related to the problem	Critical Thinking
7	I feel confident in my ability to develop a logical, consistent approach to address a real world problem related to this course	I feel confident in my ability to develop a logical, consistent approach to address a real world problem related to this course	2	Formulate an inquiry approach driven by questions relevant to the problem	Critical Thinking
8	I can list many potential ethical issues for real world problems related to this course	I can list many potential ethical issues for real world problems related to this course	2	Recognize potential ethical issues related to addressing the problem	Ethical Reasoning
9	I can draw conclusions from data that has been collected	I can draw conclusions from data collected through this experience	2	Employ the selected inquiry approach • Collect and analyze data • Draw conclusions/ inferences (interpret)	Inquiry and Analysis
10	I am able to identify and apply information from this course to address and potentially improve real-world problem(s)	I am able to identify and apply information from this course to address and potentially improve real world problem(s)	2	Apply findings toward addressing the problem	Global Learning

Item No.	Pre-Experience Survey Items	Post-Experience Survey Items	SLO	VALUE Rubrics Items	VALUE Rubric
11	I am often told I listen to and respect the ideas of others	My classmates would say that I often listened to and respected the ideas of others	3	Participate in collaborative interactions; Support group Processes; Be attentive to the ideas of others	Teamwork
12	I am often told I offer relevant questions and comments within a group setting	My classmates would say that I was able to offer relevant question and comments within a group setting	3	Participate in collaborative interactions; Support group processes; Offer relevant questions and comment	Teamwork and Civic Engagement
13	I meet obligations for group assignments on a timely basis	I met obligations for group assignments on a timely basis	3	Support group processes; Meet obligations for group assignments on a timely basis	Teamwork
14	In the past, I have purposefully reflected on what I learned from problems I encountered during a learning experience	I purposefully reflected on what I learned from problems I encountered during this experience	4	Use structured reflection in assessing an engaged inquiry experience; Use reflection on the inquiry process to guide lifelong learning	Integrative Learning
15	In the past, I often reflected on what I have learned about myself from learning experiences	During this experience, I reflected on what I have learned about myself from this experience	4	Assess what they have learned about themselves as an individual (self-awareness) from experiences; Use reflection on the inquiry process to guide lifelong learning	Integrative Learning
16	I have thought about what it means to be a member of the broader community	During this experience, I thought about what it means to be a member of the broader community	4	Assess what they have learned about themselves as members of the broader community	Integrative Learning

We also examined the extent to which longitudinal measurement invariance holds between the two surveys to examine growth in the Experience Learning SLOs over time. A key aspect of measuring students' growth over time hinges on the assumption that the instruments represent the same construct in the same metric over time (i.e., longitudinal measurement invariance). Findings from the analysis and the subsequent discussion will provide insight into the quality of the surveys as measures of the Experience Learning SLOs. Moreover, the results will describe the relationship between the Experience Learning SLOs and the caliber in which the surveys can indirectly measure student growth.

Methods

Study Sample

Two surveys were completed by different cohorts of students enrolled in courses that were redesigned to incorporate experiential learning as the main pedagogy and dispersed across five semesters (Fall 2017 through Fall 2019). The first survey, the pre-experience survey, was administered at the beginning of each semester, while the second survey, the post-experience survey, was administered at the end of each semester. All students were exposed to lifelong learning, application of knowledge and skills, collaboration with others, and structured reflection, regardless of the experiential learning course platform (e.g., internship, service learning, simulation/gaming/role-playing, study abroad, and undergraduate research); therefore, all items were deemed relevant to all survey respondents. Of the 990 students who completed at least one survey (78.7% response rate), 858 students completed the pre-experience student survey (68.2% response rate), 683 students completed the post-experience student survey (54.3% response rate), and 551 students completed both surveys (43.8% response rate).

Measurement Instruments

A key aspect of measuring students' growth over time hinges on the assumption that the instruments represent the same construct in the same metric over time (i.e., longitudinal measurement invariance).

The pre-experience and post-experience surveys measure student perceptions of achievement of the SLOs as a form of indirect assessment. The surveys were developed to provide supportive evidence for the institutional continuous improvement initiative, Experience Learning, as required by the regional accreditation agency, the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC). The pre-experience survey serves as a baseline measure, whereas the post-experience survey measures perceived learning after the experience and is compared to the pre-experience survey with the intention to measure growth. The SLOs represent the culmination of Experience Learning, a program that seeks to enhance student learning in four particular areas: lifelong learning, application of knowledge and skills, collaboration, and structured reflection. The four interrelated Experience Learning SLOs are derived from the Experience Learning mission statement, which calls for "enhancing opportunities for students to learn through actual involvement with problems and needs in the larger community," and Kolb's experiential learning cycle.

The Experience Learning SLOs highlight each stage of Kolb's experiential learning cycle. SLO 2 "Students will apply knowledge, values, and skills in solving real-world problems" aligns with the "concrete experience" and "active experimentation" stages in Kolb's cycle, while SLO 4 "Students will engage in structured reflection as part of the inquiry process" is aligned with the "reflective observation" and "abstract conceptualization" stages in Kolb's cycle. SLO 1, "Students will value the importance of engaged scholarship and lifelong learning" and SLO 3 "Students will work collaboratively with others" address the criticisms of Kolb's theory (e.g., Bergsteiner & Avery, 2014) to include collaboration and lifelong learning.

In addition to the connection with Kolb's experiential learning cycle, each SLO is accompanied by a set of benchmarks that are modified from the AAC&U VALUE rubrics. The benchmarks are used to operationalize the SLOs and guide the assessment measures. The surveys serve as an indirect measure of these learning outcomes such that each item is theoretically aligned with a benchmark from each SLO, as shown in Table 1. For example, the benchmark from SLO 1, related to lifelong learning, is constructed from language found in the "Foundations and Skills for Lifelong Learning" and "Civic Engagement" rubrics. Survey items are then reconstructed from the benchmarks to use simpler language, more suitable for undergraduate students to comprehend and answer. A similar method was used to reconstruct benchmark language for survey use with SLOs 2, 3, and 4.

After the instruments were constructed, a panel of 25 experiential learning campus experts examined the surveys for content validity. Three items from each survey were suggested to be dropped or consolidated because they were considered to be redundant and to limit survey fatigue. The surveys were pilot tested with a cohort of 80 students from five experiential learning courses. Initial results through an exploratory factor analysis revealed strong evidence that items in each survey factored onto the anticipated latent trait. Only one change occurred as a result of the pilot test; the rating scale was expanded from a 5-point Likert scale to a 7-point Likert scale (i.e., strongly disagree to strongly agree) to better examine variability between responses and to mitigate a ceiling effect. Table 2 provides descriptive statistics for the final items.

This study examined construct validity and measurement invariance between the pre-experience and post-experience surveys.

Table 2

	Mean	SD	Mean	SD
SLO1: Lifelong learning				
Item 1	5.561	1.12	6.088	1.14
Item 2	6.424	0.80	6.441	0.87
Item 3	5.166	1.27	6.201	1.09
Item 4	6.128	0.94	6.167	1.13
SLO 2: Solving real-world problems				
Item 5	5.411	1.24	6.183	0.96
Item 6	5.279	1.28	6.116	1.08
Item 7	5.389	1.27	6.183	0.97
Item 8	5.388	1.29	6.110	1.06
Item 9	5.971	0.98	6.199	1.00
Item 10	5.515	1.21	6.221	1.00
SLO 3: Collaboration				
Item 11	6.045	1.02	6.517	0.71
Item 12	5.897	0.95	6.328	0.91
Item 13	6.443	0.74	6.505	0.83
SLO 4: Structured reflection				
Item 14	5.930	0.94	6.136	1.02
Item 15	5.980	0.97	5.990	1.22
Item 16	5.892	1.10	6.044	1.18

Analysis

This study examined construct validity and measurement invariance between the pre-experience and post-experience surveys. Both surveys were designed to measure the same SLOs, and therefore, the same latent factors; however, items between the pre-experience and post-experience surveys were worded differently to better articulate the students' experiences relevant to the timing of the administered surveys. Empirical differences between corresponding items will need to be examined to justify that both surveys are measuring the same latent factors (i.e., measurement invariance). While rare, measurement invariance has been examined across surveys with altered, but theoretically aligned, items to better capture experiences across time (e.g., Vianello et al., 2018; Wang et al., 2017). Confirmation of the relationship between the survey items and the Experience Learning SLOs (i.e., the latent factors) will first be examined through a confirmatory factor analysis, as will the relationship between each of the Experience Learning SLOs. A confirmatory factor analysis provides a psychometric evaluation of the latent structure of the measurement model (Brown & Moore, 2012). A variety of fit indices (i.e., scaled-MLR chi-squared test, CFI, TLI, RMSEA, and SRMR) were used to evaluate the measurement model. Specifically, we assessed model fit using the following guidelines (Brown, 2015; Gana & Broc, 2019): comparative fit index (CFI) and Tucker-Lewis index (TLI) ≥ .90 for adequate fit and ≥ .95 for good fit; root mean square error of approximation (RMSEA) ≤ .08 for adequate fit and ≤ .05 for good fit; and standardized root mean square residual (SRMR) ≤ .08 for adequate fit and ≤ .06 for good fit. Examination of correlation between latent factors within and between occasions provided insight into relationships between the SLOs.

Next, we examined longitudinal measurement invariance between corresponding indicators and latent constructs in the pre-experience and post-experience surveys. Longitudinal measurement invariance demonstrates that the same indicators consistently measure the same construct over multiple occasions (Meredith, 1993; Millsap & Olivera-Aguilar, 2012). Model fit statistics were used to analyze invariance between models. Measurement invariance between nested models is typically assessed through a chi-squared difference test; however, research has shown that the chi-squared test is sample size dependent (Brannick, 1995; Cheung & Rensvold 2002; Kelloway, 1995). Chen (2007) argued that changes in fit indices that are independent of sample size (e.g., CFI and RMSEA) between nested models should be used to test measurement invariance rather than the chi-squared difference test. Specifically, ΔCFI $\geq$ -.010 and ΔRMSEA $\geq$ .015 indicate non-invariance (Chen, 2007); however, values of ΔCFI should take precedent as model complexity affects ΔRMSEA.

Results

A confirmatory factor model measuring SLO 1 "lifelong learning" (four manifest variables), SLO 2 "application of knowledge and skills" (six manifest variables), SLO 3 "collaboration" (three manifest variables), and SLO 4 "structured reflection" (three manifest variables) was examined using Mplus v. 8.1 (Muthén & Muthén, 1998-2017). Linearity was assessed by bivariate scatterplots and correlations among indicators. Violations of univariate and multivariate normality were prevalent in the sample. The violation of normality was addressed using a robust estimation method, robust maximum likelihood (MLR), as recommended by Lei and Shiverdecker (2019).

First, a configural invariance model was specified to examine the factor structure (i.e., whether the item measure the experience learning SLOs) separately for the pre- and post-experience surveys. This model included four correlated factors across two occasions (i.e., pre-experience survey and post-experience survey), such that eight latent factors were allowed to correlate without constraint. Correlated factors were estimated simultaneously with factor means fixed to 0 and factor variances fixed to 1 for identification. Factor loadings and intercepts were allowed to freely estimate. Residual covariances between the corresponding indicators across occasions were also freely estimated. Figure 1 outlines the standardized parameters and identification decisions for the model. The model displayed good fit, $\chi^2_{MLR}(420) = 843.65$, $p < .001$, CFI = .957, TLI = .949, RMSEA = .032, 90% CI_{RMSEA} (.029, .035), SRMR = .048. Statistically significant correlations at each occasion were found between all latent factors, as shown in Figure 1 and Table 3. Specifically, correlations between the "lifelong learning" latent factors were .354, "application of knowledge and skills" factors were .272 "collaboration" factors were .293, and "structured reflection" factors were .339. Parameter constraints were then applied to subsequent models (i.e., metric and scalar invariance models) to examine measurement invariance between two occasions.

Examination of correlation between latent factors within and between occasions provided insight into relationships between the SLOs.

Unstandardized factor loadings between corresponding indicators were constrained to be equal across both occasions to examine metric invariance. Metric measurement invariance is achieved when factor loadings are equal between occasions indicating that the factors have the same meaning across occasions. Factor variance was fixed to 1 for the pre-experience factors but allowed to estimate freely for the post-experience factors. Intercepts and residual variances were allowed to vary across occasions. Factor and residual covariances continued to be estimated freely between corresponding factors and indicators, respectively. Although, the metric model fit worse than the configural model[1] , $\Delta\chi^2_{MLR}(12) = 124.67$, $p < .001$, ΔCFI = -0.015, ΔRMSEA = 0.005, ΔRMSEA was acceptable; moreover, the model fit indices for the full metric model were within acceptable ranges, CFI = .942, TLI = .933, RMSEA = .037, 90% CI_{RMSEA} (.034, .040), SRMR = .072. The Lagrange Multiplier Test suggested item 9 as a source

[1] It should be noted that the metric model will inherently fit worse than the configural model; instead, we measured the extent to which the metric model fits worse than the configural model with the desire that the two models do not statistically differ from one another.

Figure 1.
Configural invariance model with fixed factor variance and standardized parameters

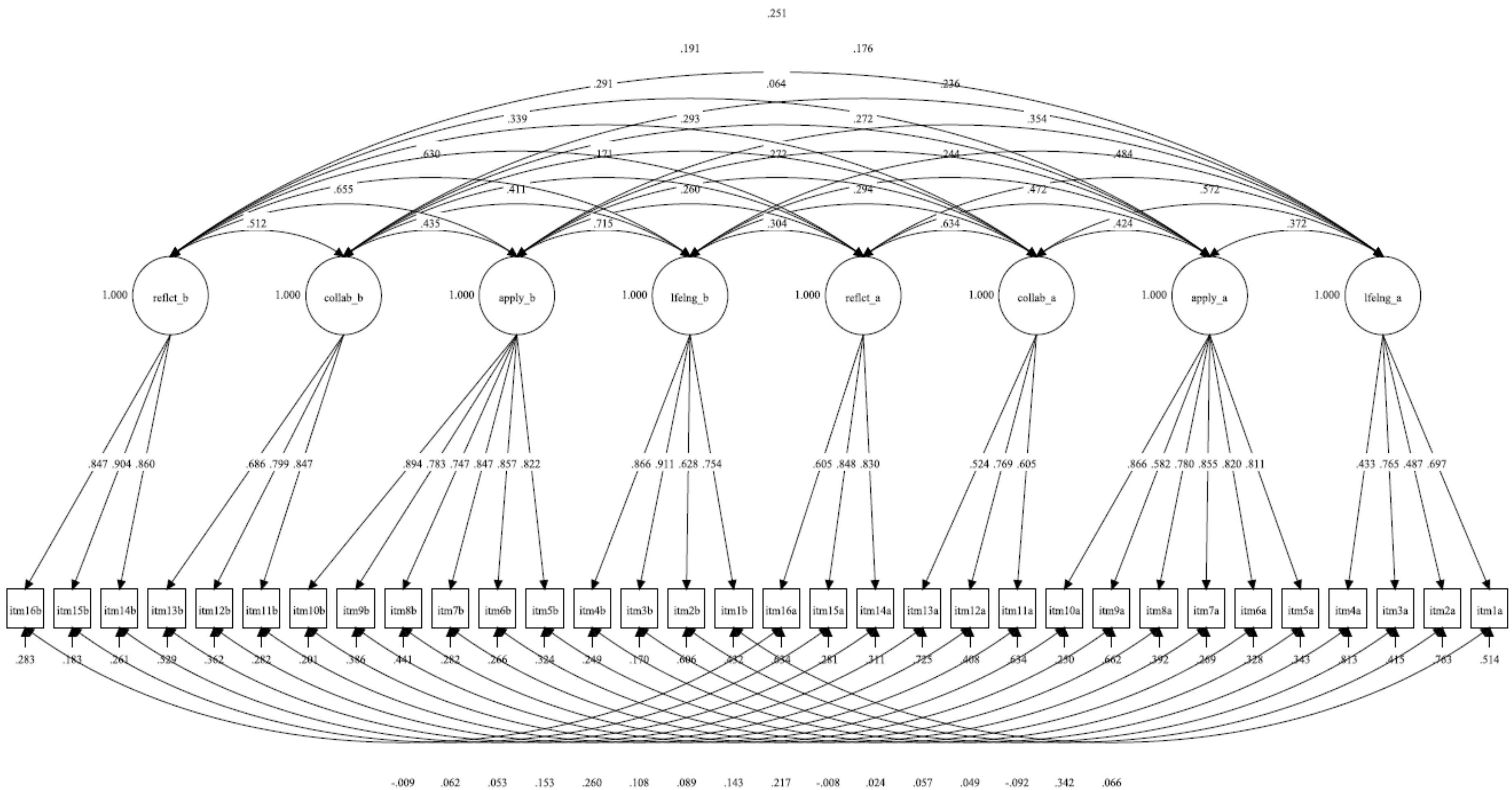

Table 3
Standardized factor correlations in configural model

	Pre-Experience Survey				Post-Experience Survey			
	SLO 1	SLO 2	SLO 3	SLO 4	SLO 1	SLO 2	SLO 3	SLO 4
Pre-Experience Survey								
SLO 1: Lifelong learning	1.00***							
SLO 2: Application of knowledge & skills	.372***	1.00***						
SLO 3: Collaboration	.572***	.424***	1.00***					
SLO 4: Structured reflection	.484***	.472***	.634***	1.00***				
Post-Experience Survey								
SLO 1: Lifelong learning	.354***	.244***	.294***	.304***	1.00***			
SLO 2: Application of knowledge & skills	.236***	.272***	.272***	.260***	.715***	1.00***		
SLO 3: Collaboration	.176**	.064	.293***	.171**	.411***	.435***	1.00***	
SLO 4: Structured reflection	.251***	.191***	.291***	.339***	.630***	.655***	.512***	1.00***

***$p < .001$, **$p < .01$

Differences between corresponding latent means between the pre-experience and post-experience surveys indicate significant student growth when post-experience survey latent means were allowed to vary, except in SLO 1.

of misfit. Modification to the model allowed factor loadings for item 9 to freely vary across occasions to produce a partial metric model[2]. The partial metric model was compared to the configural model and found to be invariant, $\Delta\chi^2_{MLR}(11) = 78.61$, $p < .001$, ΔCFI = -0.009, ΔRMSEA = 0.003. Table 4 outlines the model comparisons between the full and partial metric model to the configural model. Partial metric invariance held, which shows that the same latent factors were being measured at both occasions when item 9 was allowed to vary. Next, intercepts and factor loadings between corresponding indicators were constrained equal across both occasions to examine scalar invariance, except for item 9 which was allowed to vary freely. Factor variance and means were fixed to 1 and 0, respectively, for the pre-experience factors to allow for identification. Factor variance and means for the post-experience factors were permitted to freely estimate, while residual variances were still allowed to vary across occasions. Factor and residual covariances continued to be estimated freely between corresponding factors and indicators, respectively. The full scalar invariance model fit worse than the partial metric invariance model , $\Delta\chi^2_{MLR}(11) = 435.23$, $p < .001$, ΔCFI = -0.036, ΔRMSEA = 0.009, particularly in regards to the ΔCFI. The Lagrange Multiplier Test suggested that the intercept of item 3 to be a source of misfit. Modification to the model allowed the intercepts between occasions to freely vary. The partial scalar model was compared to the partial metric model and found to still be a worse fit, $\Delta\chi^2_{MLR}$ $(10) = 186.97$, $p < .001$, ΔCFI = -0.015, ΔRMSEA = 0.004. The Lagrange Multiplier Test suggested that the intercept of item 1 to be the next largest remaining source of misfit and was allowed to freely vary. After doing so, the new partial scalar model was found to fit the model similarly to the partial metric model, $\Delta\chi^2_{MLR}$ $(9) = 96.57$, $p < .001$, ΔCFI = -0.008, ΔRMSEA = 0.002. Table 4 outlines the model comparisons between the full scalar model and each modified scalar model to the partial metric model.

Scalar invariance holds across 13 of the 16 items which indicates that the observed differences in these indicator means between the pre-experience and post-experience surveys are due to factor mean differences only; however, items 1 and 3 had a lower expected indicator response at the same absolute level of the "lifelong learning" factor in the pre-experience survey than the post-experience survey. Item 9 had higher expected responses in the pre-experience survey than the post-experience survey in the "application of knowledge and skills" latent factor. Differences in indicator intercepts between occasions suggests that precautions should be considered when comparing factor mean differences across occasions for the "lifelong learning" latent factor. Differences between corresponding latent means between the pre-experience and post-experience surveys indicate significant student growth when post-experience survey latent means were allowed to vary, except in SLO 1, $\Delta\gamma = .087$, S.E. = .066,

[2] It should be noted that the scalar model will inherently fit worse than the metric model; instead, we measured the extent to which the scalar model fits worse than the metric model with the desire that the two models do not statistically differ from one another.

Table 4
Model comparisons for validation of longitudinal measurement invariance

Model	χ^2_{MLR}	*df*	Scaling	CFI	TLI	RMSEA	SRMR	BIC	$\Delta\chi^2_{MLR}$	Δdf	ΔCFI	ΔRMSEA
Configural Model	843.65	420	1.272	.957	.949	.032	.048	59054.5				
Metric Models[a]												
Full metric	1006.92	432	1.286	.942	.933	.037	.072	59193.4	124.67***	12	-.015	.005
Partial metric (item 9)	944.05	431	1.285	.948	.940	.035	.059	59118.5	78.61***	11	-.009	.003
Scalar Models[b]												
Full scalar (item 9)	1303.71	442	1.279	.912	.902	.044	.079	59603.8	435.23***	11	-.036	.009
Partial scalar (items 9, 3)	1103.23	441	1.280	.933	.924	.039	.069	59249.3	186.97***	10	-.015	.004
Partial scalar (items 9, 3, 1)	1029.13	440	1.281	.940	.933	.037	.064	59161.7	96.57***	9	-.008	.002

Note. χ^2_{MLR}, scaled robust maximum likelihood chi-square test; Scaling, scaling correction factor for chi-squared test for MLR estimator; CFI, comparative fit index; TLI, Tucker-Lewis index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; BIC, Bayesian information criterion; $\Delta\chi^2_{MLR}$, scaled chi-squared difference test for MLR estimator; ΔCFI, change in CFI; ΔRMESA, change in RMESA.
[a]factor loadings in items listed in parenthesis are free to vary between occasions.
[b]intercepts in items listed in parenthesis are free to vary between occasions.

p = .188. There exists a significant difference between the latent means of pre-post SLO 2, $\Delta\gamma$ = .742, S.E. = .041, p < .001; pre-post SLO 3, $\Delta\gamma$ = .611, S.E. = .054, p < .001; and pre-post SLO 4, $\Delta\gamma$ = .163, S.E. = .058, p = .005.

In addition to assessing the construct validity and longitudinal measurement invariance, we also calculated the average variance extracted (AVE) from each factor and the internal consistency reliability for each factor. The AVE is useful in assessing the presence of convergent and divergent validity in the model. Convergent and divergent validity help supplement the claim for construct validity by determining the degree to which the factors are related and unrelated to one another, respectively. Cheung and Wang (2017) recommended that convergent validity be established provided that AVE and standardized factor loadings of all items are not considerably less than .500. The AVE was greater than .500 for six out of the eight factors. Two factors in the pre-experience survey were slightly lower than this threshold: "lifelong learning" AVE = .430 and "collaboration" factor was .428. The AVE values for each factor are presented in Table 4. Additionally, standardized factor loadings for all but two items were above the .500 threshold: item 2 (λ = .433) and item 4 (λ = .487) in the pre-experience survey. These results suggest strong convergent validity for the post-experience survey. While the instruments overall exhibited strong factor loadings and AVE, the "lifelong learning" and "collaboration" factors in the pre-experience survey revealed values that were slightly below the threshold to conclude convergent validity for the pre-experience survey. Divergent validity is concluded if the correlation between any two factors is not considerably greater than .700 (Cheung & Wang, 2017). Table 3 shows correlations between factors to be below the threshold with the exception of the "lifelong learning" and "application of knowledge and skills" factors in the post-experience survey, which has a correlation of .715. Thus, divergent validity is evident within the model.

These results suggest strong convergent validity for the post-experience survey.

Internal consistency was measured using McDonald's (1999) ω_t coefficient. The ω_t coefficient, along with a confidence interval, provides a better reflection of variability within the point estimation process than the τ-equivalent model (i.e., Cronbach's alpha); that is, ω_t reflects a more accurate degree of confidence in the internal consistency of the factor (Dunn et al., 2013). Confidence intervals (95%) were obtained through the bias-corrected and accelerated bootstrapping technique in the MBESS package 4.6.0 in R 3.5.1. Table 5 outlines the ω_t internal consistency coefficient for the model factors. Factors display strong internal consistency with all but one factor above ω_t = .700. The "collaboration" factor in the pre-experience survey showed moderate levels of internal consistency, ω_t = .682, 95% CI (.632, .722).

Table 5
Average variance extracted (AVE); composite reliability (CR); and ω_t internal consistency estimates, standard errors, and confidence intervals for model factor

					ω_t CI 95%	
	AVE	CR	ω_t	S.E.	Lower	Upper
Pre-Experience Survey						
SLO 1: Lifelong learning	.430	.694	.730	.018	.691	.761
SLO 2: Solving real-world problems	.644	.908	.913	.005	.903	.922
SLO 3: Collaboration	.428	.671	.682	.023	.632	.722
SLO 4: Structured reflection	.571	.810	.794	.015	.759	.821
Post-Experience Survey						
SLO 1: Lifelong learning	.644	.873	.883	.013	.856	.908
SLO 2: Solving real-world problems	.681	.928	.927	.008	.912	.944
SLO 3: Collaboration	.602	.823	.815	.027	.758	.866
SLO 4: Structured reflection	.761	.904	.905	.010	.883	.921

Note: Bias-corrected and accelerated bootstrap confidence interval used with 1000 bootstrap iterations.

Discussion

Our findings provide support for the instruments to measure the desired Experience Learning SLOs, and importantly that this structure is mostly invariant over time (i.e., equality of factor structure, factor loadings, and intercepts).

Kolb's experiential learning model focuses on the active (application of knowledge and skills) and transformative (structured reflection) aspects of learning. As part of the survey construction, items related to these facets of the model were carefully constructed based on the language used in the critical thinking, ethical reasoning, inquiry and analysis, global learning, and integrative learning AAC&U VALUE rubrics and the principles laid out in Kolb's experiential learning theory (Kolb, 1984). We set out to develop pre-experience and post-experience surveys to measure students' attainment of specific Experience Learning SLOs. A key aspect of measuring students' growth over time hinges on the assumption of longitudinal measurement invariance (i.e., the instruments need to represent the same construct in the same metric over time). Thus, the purpose of the present study was to test the construct validity and longitudinal measurement invariance of the pre-experience and post-experience surveys.

Our findings provide support for the instruments to measure the desired Experience Learning SLOs, and importantly that this structure is mostly invariant over time (i.e., equality of factor structure, factor loadings, and intercepts). Our results overall indicate good psychometric properties of the instruments to measure SLOs in experiential learning courses and demonstrate that meaningful comparisons can be made to assess students' growth on these learning outcomes. Although we did not achieve full metric or scalar invariance, partial scalar invariance was achieved for the vast majority of items (i.e., 13 out of 16 items). According to Sass and Schmitt (2013), "if the number non-invariant items is small compared to total number of items, or the overall amounts of non-invariance is small, the latent factor means used for group comparisons should not be drastically impacted" (p. 324). Knowing which specific items are non-invariant, as evidenced in our sample, researchers and assessment professional applying these instruments can modify the problematic items in the future with the aim of achieving full metric and scalar invariance.

Our results demonstrate at least partial scalar longitudinal measurement invariance between the two instruments over the course of a semester suggesting that the surveys measured the same SLOs at different occasions. This also implies that the mean differences (or student growth) can be interpreted as true changes in students' attainment of the SLOs. These findings have significant implications for assessment professionals examining learning outcomes in experiential learning settings. In particular, attainment and growth in learning outcomes commonly found in experiential learning contexts can be measured, at least from the students' perspective. These results, in conjunction with valid direct measures, provide strong evidence of student attainment over the course of an experiential course or other learning experience (Banta & Palomba, 2015).

Convergent validity occurs when the degree to which items under the same construct are related to one another. Measures of AVE and standardized factor loadings greater than

.500 were used to support the claim for convergent validity (Cheung & Wang, 2017). While two factors in the pre-experience survey (i.e., "lifelong learning" and "collaboration") failed to meet the AVE threshold and two items in the pre-experience "lifelong learning" factor failed to meet the standardized factor loading threshold to support convergent validity, Fornell and Larcker (1981) argue that AVE values greater than .400 are adequate to support convergent validity, provided that composite reliability (CR) values are at least .600. The two factors whose AVE < .500 were still above the .400 threshold and had CR values exceeding .600 (see Table 4); thus, evidence for convergent validity is arguable. The internal consistency coefficient (ω_t) also supports that the items measure their respective factors. The degree to which the surveys discriminate between the constructs within the model (i.e., divergent validity) is supported through examination of the correlations between factors. Cheung and Wang (2017) posited that divergent validity is evident when any correlation between two factors is not significantly greater than .700. The factor correlation between post-experience "lifelong learning" and "applications of knowledge and skills" did exceed .700, but not substantially so (i.e., r = .715); moreover, Kline (2010) suggested a more liberal cut-off of .850, which may be more reasonable considering the theorized relationship between all factors (Kolb & Kolb, 2011).

When the surveys are used jointly, they can quantify the practical impact of the experiential learning environment from the learners' perspective.

Correlations between corresponding factors (e.g., pre-experience survey "lifelong learning" and post-experience survey "lifelong learning") exhibited the highest coefficients compared to non-corresponding factors between occasions; that is, the correlation between the pre-experience survey "lifelong learning" factor was strongest with the post-experience "lifelong learning" factor than any other post-experience factor. The same results were found among all corresponding factors between surveys. These results indicate that the differences between occasions is mostly to do with the change across the same SLO. Additionally, strong internal consistency reliability within each factor was evident and suggests that items were likely reliably measuring their latent factors.

The pattern of correlations among the factors both within and between occasions indicates that our SLOs exhibit non-cyclical relationships between one another. Warren et al. (1995) suggested that experiential learning objectives are not achieved in any particular order or pattern. Kolb's (1984) model shows that learning in an experiential context is patterned such that lifelong learning would follow reflection, reflection would follow applications, and so forth. Although this relationship still holds in our model, one SLO does not necessarily precede a specific SLO but rather, any factor could precede or succeed another and provide substantive evidence for student growth across all SLOs.

The pre-experience and post-experience surveys are intended to be used to understand student attainment of learning outcomes commonly found in an experiential learning opportunity, such as study abroad, service-learning, internships, undergraduate research, and others. Intended users should implement the surveys before and immediately after an experiential activity to better measure student growth as a result of the real-world learning experience. Survey results can be examined by individual items and many can be compared across multiple occasions with exceptions to a few items. In particular, items in SLO 3 "collaboration" and SLO 4 "structured reflection" are invariant over time. Users are encouraged to replicate the study with different samples and adjusted language to validate and improve our findings. When the surveys are used jointly, they can quantify the practical impact of the experiential learning environment from the learners' perspective. The learners' perspective is key when they are the most valuable and reliable source to understand the role of the experiential learning such as during internships, externships, practicums, and study abroad experiences. In particular, the surveys serve as strong indirect measures of the Experience Learning SLOs and Kolb's experiential learning cycle.

References

Abdulwahed, M., & Nagy, Z.K. (2009). Applying Kolb's experiential learning cycle for laboratory education. *Journal of Engineering Education*, *98*(3), 283-294. https://doi.org/10.1002/j.2168-9830.2009.tb01025.x

Baker, M.A., Robinson, J.S., & Kolb, D.A. (2012). Aligning Kolb's experiential learning theory with comprehensive agricultural education model. *Journal of Agricultural Education*, *53*(4), 1-16. https://doi.org/10.5032/jae.2012.04001

Banta, T.W., & Palomba, C.A. (2015). *Assessment Essentials: Planning, implementing, and improving assessment in higher education* (2nd ed.). Jossey-Boss.

Bergsteiner, H., & Avery, G.C. (2014). The twin-cycle experiential learning model: Reconceptualising Kolb's theory. *Studies in Continuing Education*, *36*(3), 257-274. https://doi.org/10.1080/0158037X.2014.904782

Bergsteiner, H., Avery, G.C., & Neumann, R. (2010). Kolb's experiential learning model: critique from a modelling perspective. *Studies in Continuing Education*, *32*(1), 29-46. https://doi.org/10.1080/01580370903534355

Boyatzis, R.E., Cowen, S.S., & Kolb, D.A. (1995). *Innovation in professional education: Steps on a journey from teaching to learning*. Jossey-Boss.

Brannick, M.T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*(3), 201-213. https://doi.org/10.1002/job.4030160303

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Brown, T.A., & Moore, M.T. (2012). Confirmatory factor analysis. In R.H. Hoyle. (Ed.), *Handbook of structural equation modeling*, (pp. 361-379). Guilford Press.

Chapman, S., McPhee, P., & Proudman, B. (1992). What is experiential education? *Journal of Experiential Education*, *15*(2), 16-23. https://doi.org/10.1177/105382599201500203

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cheung, G.W., & Wang, C. (2017). Current approaches for assessing convergent and discriminant validity with SEM: Issues and solutions. [Abstract]. *Academy of Management*, *1*. https://doi.org/10.5465/ AMBPP.2017.12706abstract

Dewey, J. (1938). *Experience and education*. MacMillan.

Dunn, T.J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *The British Journal of Psychology*, *105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Finley, A. (2011). How reliable are the VALUE rubrics? *Peer Review: Emerging trends and key debates in undergraduate education*, *14*(4), 31-34. https://www.aacu.org/sites/default/files/files/peerreview/PRfa11wi12.pdf

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement errors. *Journal of Marketing Research*, *18*(1), 39-80. https://doi.org/10.2307/3151312

Gana, K., & Broc, G. (2019). *Structural equation modeling with Lavaan*. Wiley.

Healey, M., & Jenkins, A. (2000). Kolb's experiential learning theory and its application in geography in higher education. *Journal of Geography*, *99*(5), 185-195. https://doi.org/10.1080/00221340008978967.

Kayes, A.B., Kayes, C., & Kolb, D.A. (2005). Experiential learning in teams. *Simulation & Gaming*, *36*(3), 330-354. https://doi.org/10.1177/1046878105279012

Kelloway, E.K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior*, *16*(3), 215-224. https://doi.org/10.1002/job.4030160304

Kline, R.B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Publications.

Kolb, D.A. (1984). *Experiential learning: Experience as a course of learning and development*. Prentice-Hall.

Kolb, Y.A., & Kolb, D.A. (2011). Experiential learning theory: A dynamic, holistic approach to management learning, education and development. In S.J. Armstrong, & C.V. Fukami (Eds.), *The SAGE handbook of management learning*, *education and development*, (pp. 42-68). SAGE Publications. https://doi.org/10.4135/9780857021038.n3

Lei, P., & Shiverdecker, L.K. (2019). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(4), 584-601. https://doi.org/10.1080/10705511.2019.1680292

Manolis, C., Burns, D.J., Assudani, R., & Chinta, R. (2013). Assessing experiential learning styles: A methodological reconstruction and validation of the Kolb learning style inventory. *Learning and Individual Differences*, *23*(1), 44-52. https://doi.org/10.1016/j.lindif.2012.10.009

McDonald, R.P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates. https://doi org/10.1111/j.2044-8317.1981.tb00621.x

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543. https://doi.org/10.1007/BF02294825

Miettinen, R., (2010). The concept of experiential learning and John Dewey's theory of reflective thought and action. *International Journal of Lifelong Education*, *19*(1), 54-72. https://doi.org/10.1080/026013700293458

Millsap, R.E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R.H. Hoyle. (Ed.), *Handbook of structural equation modeling*, (pp. 380-392). Guilford Press.

Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus user's guide* (8th ed.). Author.

Poore, J.A., Cullen, D.L., & Schaar, G.L. (2014). Simulation-based interprofessional education guided by Kolb's experiential learning theory. *Clinical Simulation in Nursing*, *10*(5), 241-247. https://doi.org/10.1016/j.ecns.2014.01.004

Petkus, E. (2000). A theoretical and practical framework for service-learning in marketing: Kolb's experiential learning cycle. *Journal of Marketing Education*, *22*(1), 64-70. https://doi.org/10.1177/0273475300221008

Reshmad'sa, L., & Vijayakumari, S.N. (2017). Effect of Kolb's experiential learning strategy on enhancing pedagogical skills of pre-service teachers in secondary school level. *Journal of School Educational Technology*, *13*(2), 1-6. https://doi.org/10.26634/jsch.13.2.13825

Rhodes, T., & Finley, A. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Association of American Colleges and Universities. https://www.aacu.org/publications/using-value-rubrics

Rhodes, T. (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Association of American Colleges and Universities. https://www.aacu.org/publications-research/publications/assessing-outcomes-and-improving-achievement-tips-and-tools-using

Russell-Bowie, D. (2013). Mission impossible or possible mission? Changing confidence and attitudes of primary preservice music education students using Kolb's experiential learning theory. *Australian Journal of Music Education*, *2*, 46-63. https://files.eric.ed.gov/fulltext/EJ1061833.pdf

Sass, D.A., & Schmitt, T.A. (2013). Testing measurement and structural invariance: Implications for practice. In T. Teo (Ed.), *Handbook of quantitative methods for educational research*, (pp. 315-345). Sense Publishers.

Seaman, J., Brown, M., & Quay, J. (2017). The evolution of experiential learning theory: Tracing lines of research in the JEE. *Journal of Experiential Education*, *40*(4), https://doi.org/10.1177/1053825916689268

Smith, K.L., & Rayfield, J. (2017). Student teaching changed me: A look at Kolb's learning style inventory scores before and after the student teaching experience. *Journal of Agricultural Education*, *58*(1), 102-117. https://doi.org/10.5032/jae.2017.01102

Vianello, M., Dalla Rosa, A., Anselmi, P., & Galliani, E.M. (2018). Validity and measurement invariance of the unified multidimensional calling scale (UMCS): A three-wave survey study. *PLoS ONE*, *13*(12). https://doi.org/10.1371/journal.pone.0209348

Wang, M., Chen, R.C., Usinger, D.S., & Reeve, B.B. (2017). Evaluating measurement invariance across assessment modes of phone interview and computer self-administrated survey for the PROMIS measures in a population-based cohort of localized prostate cancer survivors. *Quality of Life Research*, *26*, 2973-2985. https://doi.org/10.1007/s11136-017-1640-3

Warren, K., Sakofs, M., & Hunt, J.S. (1995). *The theory of experiential education*. Kendall Hunt.

Widaman, K.F., Ferrer, E., & Conger, R.D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10-18. https://doi.org/10.1111/j.1750-8606.2009.00110.x