A Construct Revalidation of the Community of Inquiry Survey: Empirical Evidence for a General Factor Under a Bifactor Structure

Hongwei Yang, Ph.D.[1] and Jian Su, Ph.D.[2]
[1]University of West Florida, [2]University of Tennessee

Abstract

The study revisited the community of inquiry (CoI) instrument for construct revalidation. To that end, the study used confirmatory factor analysis (CFA) to examine four competing models (unidimensional, correlated-factor, second-order factor, and bifactor models) on model fit statistics computed using parameter estimates from a statistical estimator for ordinal categorical data. The CFA identified as the optimal structure the bifactor model where all items loaded on their intended domains and the existence of the general factor was supported, essentially evidence of construct validity for the instrument. The study further examined the bifactor model using mostly model-based reliability measures. The findings confirmed the contributions of the general factor to the reliability of instrument scores. The study concluded with validity and reliability evidence for the bifactor model, supported the model as a valid and reliable representation of the CoI instrument and a fuller representation of the CoI theoretical framework, and recommended its use in CoI-related research and practice in online education.

Keywords: community of inquiry, teaching presence, social presence, cognitive presence, bifactor model, construct revalidation

Introduction

The CoI Theoretical Framework

The community of inquiry (CoI) theoretical framework was first laid out by Garrison et al. (2000). The framework identified elements critical for understanding the dynamics of an online learning experience and for structuring and supporting the process of online teaching and learning as well as related research (Kozan & Caskurlu, 2018; Olpak & Kiliç Çakmak, 2016; Shea & Bidjerano, 2009).

The CoI framework consisted of three interconnected constructs of collaborative constructivist learning: (a) teaching presence (TP); (b) social presence (SP); and (c) cognitive presence (CP). Here, the term presence referred to fidelity: how real were the learning and the environment where it occurred (Dempsey & Zhang, 2019). The greater the presence, the greater the fidelity, and accordingly the more realistic the learning experience was perceived to be. Each presence overlapped with the other two and all three combined within a community of inquiry (Diaz et al., 2010; Kovanović et al., 2018; Kozan, 2016; Shea & Bidjerano, 2009). Finally, each presence was further conceptualized to be multidimensional with multiple categories represented by multiple indicators (Caskurlu, 2018; Garrison et al., 2000; Olpak & Kiliç Çakmak, 2016).

Cognitive Presence

Cognitive presence was described as a developmental model articulating the dynamics of a worthwhile educational experience (Garrison et al., 2010). It referred to the extent to which students in a community of inquiry were able to construct meaning through sustained communication and reflected the process of inquiry and learning (Bangert, 2009; Garrison et al., 2000). When operationalizing CP, Garrison et al. (2000) used the practical inquiry (PI) model reflecting the critical thinking process for creating CP (Olpak & Kiliç Çakmak, 2016). They expanded the PI model into a cycle of four phases (subconstructs) in the inquiry process which the CoI framework subsumed as categories: (a) triggering event, (b) exploration, (c) integration, and (d) resolution.

Social Presence

Social presence has focused on important issues that mold social climate in the online learning community and on the level of recognition (e.g., ability of learners to identify with the community, purposeful conversation in a trusting environment, development of interpersonal relationships) among learners during the process of communication (Garrison & Arbaugh, 2007; Kovanović et al., 2018). SP played an important role in creating an online learning environment that encouraged critical thinking (Bangert, 2009; Garrison et al., 2000). Studies have consistently shown that SP had a strong influence on students' satisfaction with online courses and with the instructor, and their perception of learning in online courses (Caskurlu, 2018). Finally, SP consisted of three subconstructs that the CoI framework subsumed as categories: (a) affective expression, (b) open communication, and (c) group cohesion.

Teaching Presence

Teaching presence referred to designing, facilitating, and directing of cognitive and social processes by the instructor to create meaningful personal learning and valuable learning outputs, and also focused on learner ratings of those actions by the instructor (Olpak & Kiliç Çakmak, 2016; Shea & Bidjerano, 2009).

There has been a growing recognition of the importance of TP for successful online teaching and learning, especially when critical thinking and discourse were required (Garrison et al., 2000). A community of inquiry provided important support to critical thinking and meaningful learning. After combining all elements of this community, TP should have served to facilitate critical discussion and learning in this environment. Finally, TP consisted of three subconstructs that the CoI framework subsumed as categories: (a) design and organization, (b) facilitating discourse, and (c) direct instruction.

The CoI Instrument and its Validation

The CoI framework was operationalized by Arbaugh et al. (2008) through developing and validating a CoI instrument which consisted of three subscales, each of which addressed one of the three presence constructs/factors in a domain. Each CoI item measured the extent to which an online course characteristic was present. Their results supported the instrument as being a valid, reliable, and efficient measure of the CoI framework. In their study, the Cronbach's α for TP was .94, that for SP .91, and that for CP .95.

Since the development of the CoI instrument, its refinement has been constantly called for and therefore continuous over the past 10 plus years in various online settings (Arbaugh, et al., 2008; Dempsey & Zhang, 2019; Kozan & Caskurlu, 2018; Kozan & Richardson, 2014). In many refinement studies, the original correlated, three-factor structure of the instrument was largely recovered and revalidated through either exploratory, confirmatory, or both statistical methods (Kovanović et al., 2018).

Exploratory Approach

Under an exploratory approach, the validation study typically implemented either an exploratory factor analysis (EFA) or a principal component analysis (PCA) with an oblique rotation to allow the extracted factors measuring latent constructs to be correlated. Among such studies were Swan et al. (2008), Shea and Bidjerano (2009), Díaz et al. (2010), Garrison et al. (2010), and Kovanović et al. (2018).

Swan et al. (2008) and Garrison et al. (2010) both conducted a PCA of the CoI responses ($n = 287$ master's and doctoral students and $n = 205$ master's students, respectively) with an oblimin rotation; the results supported a three-factor solution congruent with Garrison et al. (2000). Similar results were obtained in Shea and Bidjerano (2009) with $n = 2,159$ online students. Instead of a PCA, they used an EFA under principal axis factoring with an oblimin rotation.

Díaz et al. (2010) used an enhanced version of the CoI instrument that provided a second way of rating each item and conducted a PCA with an oblimin rotation of the data from 412 students. The enhanced instrument evaluated both the extent to which an online course characteristic existed (i.e., the first rating responses to the original CoI item statements) and the importance of that characteristic, as based on the second rating responses. A new score was created by multiplying the two sets of ratings. After analyzing the multiplicative scores under the PCA, the three CoI factors were successfully recovered.

Finally, Kovanović et al. (2018) proposed tweaks to the original factor structure. They used an EFA under principal axis factoring with an oblimin rotation to analyze a large sample of 1,487 students in a massive open online course setting. They largely recovered the original three-factor structure. In their analysis, item

28 in the CP subscale cross-loaded on the SP subscale. Fortunately, the removal of this item had only a minor impact on the loadings of the other items and the overall model statistics.

Confirmatory Approach

Under a confirmatory approach, the validation study typically implemented a confirmatory factor analysis (CFA) to assess the original correlated-factor model. Among such studies were Caskurlu, (2018), Dempsey and Zhang (2019), Kozan (2016), and Ma et al. (2017).

Kozan (2016) applied a CFA to validate the CoI instrument using the responses from 338 participants who were mostly online master's students. Kozan used a slightly adapted version of the survey, and the CFA model that contained the three-factor structure was successfully revalidated without any re-specification.

Ma et al. (2017) added to the CoI instrument a new, learning presence proposed by Shea and Bidjerano (2010) and measured by an additional set of 14 items. The sample was 325 undergraduate students in a blended learning environment. Their findings supported the fit of the four-factor model after two rounds of CFA which led to the deletion of one item.

Caskurlu (2018) conducted a CFA to investigate the factor structure of each individual presence using a dataset of 310 participants, and established that each presence itself was multidimensional and thus a higher-order construct. The analysis was run separately for each individual subscale without examining the dimensionality of the instrument as a whole.

Dempsey and Zhang (2019) revalidated the CoI instrument using a dataset from 579 online MBA students. They experimented with multiple structures: (a) the original three-factor structure, (b) a 10-factor structure by allowing each of the three presences to be multidimensional, and (c) a higher-order factor model with three lower-order factors. The study concluded the higher-order factor model provided the best fit to the data.

Both Exploratory and Confirmatory Approaches

There are also validation studies which used both an EFA/PCA and CFA (Bangert, 2009; Kozan & Richardson, 2014; Olpak & Kiliç Çakmak, 2016; Yu & Richardson, 2015). Such studies typically had a large sample which was randomly split into two subsets with each one still being large enough for either an EFA/PCA or CFA. Then, one random subset was analyzed to explore and discover the underlying factor structure of the instrument, and the finding was next further assessed under CFA using the second random subset.

Bangert (2009) used a sample of 1,173 undergraduate and graduate students enrolled in fully online and blended courses to validate the CoI instrument. One half of the sample was randomly selected to conduct a PCA with an oblique rotation which was followed by a CFA using the remaining half of the sample. The EFA process largely recovered the original three-factor solution. Next, this recovered factor structure was confirmed by CFA with all fit statistics being satisfactory.

Kozan and Richardson (2014) collected their data from master's and doctoral students who were either fully online, or face-to-face but also taking online courses. They had 219 participants for EFA and 178

participants for CFA. During EFA, they selected the three-factor solution from the promax rotation. Next, the CFA process experimented with multiple models using the second dataset and the final model exhibited a good fit. Except for several correlated errors, this final model concurred with the CoI framework.

Yu and Richardson (2015) collected data from 995 undergraduate online students and split them into two approximately equal subsets. During the EFA (promax rotation) process, they experimented with two models before and after removing two items, and ended up selecting the second, three-factor model where each item loaded on its intended subscale under the CoI framework. In the CFA, the fit of the 32-item model was confirmed with excellent values on model fit statistics.

Olpak and Kiliç Çakmak (2016) collected the data from 1,150 students enrolled in online courses and randomly split them into two equal groups. Under EFA, they successfully recovered the original three-factor structure. Then, in the CFA, the fit of the three-factor model was assessed multiple times and was confirmed to be excellent after, per the modification indices, allowing several item error covariances to be freely estimated.

In the end, a more complete summary of research on the validation of the CoI instrument was found in Kozan and Caskurlu (2018) and Stenbom (2018) which provided systematic reviews of such studies.

Issues with Existing Validation Work

Many CoI refinement studies have shared similar limitations. A primary limitation has been that the correlated-factor model most studies have universally relied on cannot fully describe the CoI framework by Garrison et al. (2000). Furthermore, there is room for improvement in the estimation method by which the estimates of model parameters have been derived.

First, the correlated-factor model has addressed only part of the CoI framework. Although it explicitly allowed the presences to be correlated in pairs, it did not include an intersection of all three presences. Such an inadequacy was unfortunate because the literature has repeatedly emphasized the importance of the interaction of all three presences which represents an online learning or educational experience (Caskurlu, 2018; Diaz et al., 2010; Garrison et al., 2000; Garrison & Arbaugh, 2007; Garrison et al., 2010; Kozan, 2016; Kozan & Richardson, 2014; Olpak & Kiliç Çakmak, 2016; Swan & Ice, 2010).

The literature has recommended the examination of competing models when applying CFA to scale validation (Gignac & Kretzschmar, 2017; Rodriguez et al., 2016b). There are several competing models for handling multidimensional data: (a) M1: unidimensional, single-factor model; (b) M2: correlated-factor model; (c) M3: second-order factor model; and (d) M4: bifactor model (Chen et al., 2012; Gignac & Kretzschmar, 2017; Reise, 2012; Reise et al., 2010; Reise et al., 2007; Rodriguez et al., 2016a, 2016b).

Figure 1 demonstrates the implication of each competing model for the CoI framework. First, corresponding to Figure 1a which consisted of a single circle, the single-factor model consisted of one and only one latent factor underlying all CoI items to model their shared variance. The factor shed light on the general factor in the bifactor model which documented a general CoI construct measuring students' online educational experience (Garrison et al., 2000; Reise, 2012). Second, corresponding to Figure 1b which consisted of three separate circles overlapping in pairs but did not include an intersection of all three circles, both the

correlated-factor and the second-order models allowed each pair of presences to be correlated either explicitly (M2) or implicitly (M3). However, neither model had a factor underlying all items, therefore both lacked the area in the CoI framework shared by all three presences which represented an online educational experience.

Finally, corresponding to Figure 1c which consisted of three separate circles overlapping in pairs and an intersection of all three circles, the bifactor model incorporated both the overlap of each pair of presences and the intersection of all three presences into the general factor underlying all items, thus indicating the bifactor structure was more aligned with the CoI framework than were the other models.

Even though it could be of interest to further distinguish the pairwise correlations between presences from the interaction of all three presences, doing so would have caused additional complications to the bifactor model. Should any pair of presences be correlated, this would suggest the existence of additional, unmodeled general factors. Also, many statistics (e.g., model-based reliability statistics) used in this study would have been challenging to implement, and any improvement in model fit would have been offset by losses in model interpretability (Reise, 2012). Therefore, this study specified the bifactor model in the usual way consistent with the literature where the general and the presence domain factors were all uncorrelated with each other, without attempting to separate the overlap between each pair of presences from the intersection of all three presences by introducing additional pairwise correlations between presences (Chen et al., 2012; Reise, 2012; Rodriguez et al., 2016a, 2016b).

Figure 1

Alignment of the CoI Theoretical Framework and Four Competing Models

(a) M1: Single-Factor Model   (b) M2: Correlated-Factor Model   (c) M4: Bifactor Model
M3: Second-Order Model
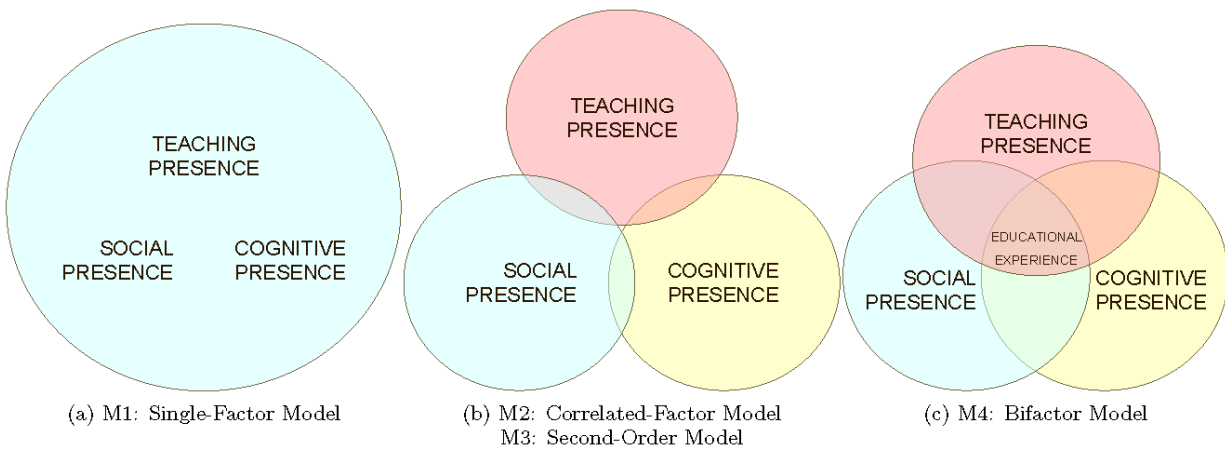
Second, previous validation studies routinely used statistical methods unable to factor into consideration the rating scale structure of the CoI item responses (i.e., ordinal categorical data). They treated the responses as if they were continuous. Such a practice has been known to have undesirable consequences: inflated χ^2 statistic, under-estimate of the standard error, and so on (Byrne, 2010; Kline, 2016). These problems were exacerbated when the number of categories was small (four/five categories or less) and/or the data exhibited serious skewness and kurtosis (outside range of -1.00 to +1.00 for skewness, -1.50 to +1.50 for kurtosis). Generally, to properly address the rating scale structure of such data as the CoI

responses, the robust weighted least squares (WLS) method and its variants have been recommended (Rosseel, 2012).

Purpose of Study

To address the inadequacies/limitations, this study began with a CFA to construct-revalidate the CoI instrument by examining the four competing models (DiStefano & Hess, 2005). After estimating the four models and identifying the one providing the optimal fit as evidence of construct validity, the study computed more statistics for the optimal structure to complement the construct-revalidation results from the CFA.

Therefore, the study proposed and addressed the following two research questions:

1. How well does each of M1 through M4 fit the CoI data as measured by commonly used model fit statistics?

2. What are the psychometric properties (e.g., validity, reliability) of the optimal model as identified above?

When addressing the two questions, the study factored into consideration the rating scale structure of the CoI responses.

Methods

The CoI survey in this study consisted of 34 five-point Likert items (see Table 1; Arbaugh et al, 2008): 1 for *strongly disagree*, 2 for *disagree*, 3 for *neutral*, 4 for *agree*, and 5 for *strongly agree*. The 34 items make up three subscales: (a) teaching presence (13 items); (b) social presence (9 items); and (c) cognitive presence (12 items).

Table 1

Community of Inquiry Survey Items

| Item	Item statement	Subscale
01	The instructor clearly communicated important course topics.	TP
02	The instructor clearly communicated important course goals.	TP
03	The instructor provided clear instructions on how to participate in course learning activities.	TP
04	The instructor clearly communicated important due dates/time frames for learning activities.	TP
05	The instructor was helpful in identifying areas of agreement and disagreement on course topics that helped me to learn.	TP

06	The instructor was helpful in guiding the class towards understanding course topics in a way that helped me clarify my thinking.	TP
07	The instructor helped to keep course participants engaged and participating in productive dialogue.	TP
08	The instructor helped keep the course participants on task in a way that helped me to learn.	TP
09	The instructor encouraged course participants to explore new concepts in this course.	TP
10	Instructor actions reinforced the development of a sense of community among course participants.	TP
11	The instructor helped to focus discussion on relevant issues in a way that helped me to learn.	TP
12	The instructor provided feedback that helped me understand my strengths and weaknesses.	TP
13	The instructor provided feedback in a timely fashion.	TP
14	Getting to know other course participants gave me a sense of belonging in the course.	SP
15	I was able to form distinct impressions of some course participants.	SP
16	Online or Web-based communication is an excellent medium for social interaction.	SP
17	I felt comfortable conversing through the online medium.	SP
18	I felt comfortable participating in the course discussions.	SP
19	I felt comfortable interacting with other course participants.	SP
20	I felt comfortable disagreeing with other course participants while still maintaining a sense of trust.	SP
21	I felt that my point of view was acknowledged by other course participants.	SP
22	Online discussions helped me to develop a sense of collaboration.	SP
23	Problems posed in this course increased my interest in course issues.	CP
24	Some course activities piqued my curiosity.	CP
25	I felt motivated to explore content-related questions.	CP
26	I utilized a variety of information sources to explore problems/issues presented in this course.	CP
27	Brainstorming and finding relevant information helped me resolve content-related questions.	CP
28	Online discussions were valuable in helping me appreciate different perspectives.	CP
29	Combining new information from a range of sources helped me answer questions raised in course activities.	CP
30	Learning activities in this course helped me construct explanations/solutions.	CP
31	Reflecting on course content and discussions helped me understand fundamental concepts in this class.	CP
32	I can describe ways to test and apply the knowledge created in this course.	CP
33	I have developed solutions to course problems that can be applied in practice.	CP
34	I can apply the knowledge created in this course to my work or other non-class related activities.	CP

Figure 2 shows the four competing models for the CoI instrument: (a) single-factor model (Figure 2a); (b) correlated-factor model (Figure 2b); (c) second-order factor model (Figure 2c); and (d) bifactor model (Figure 2d). To estimate the four models, the study used the *lavaan* package in R which offered a WLS estimator for handling ordinal categorical data (Rosseel, 2012).

Figure 2

CoI Dimensionality Analysis Under Four Competing Models Using CFA

(a) M1: Single-Factor Model

(b) M2: Correlated-Factor Model

(c) M3: Second-Order Model

(d) M4: Bifactor Model
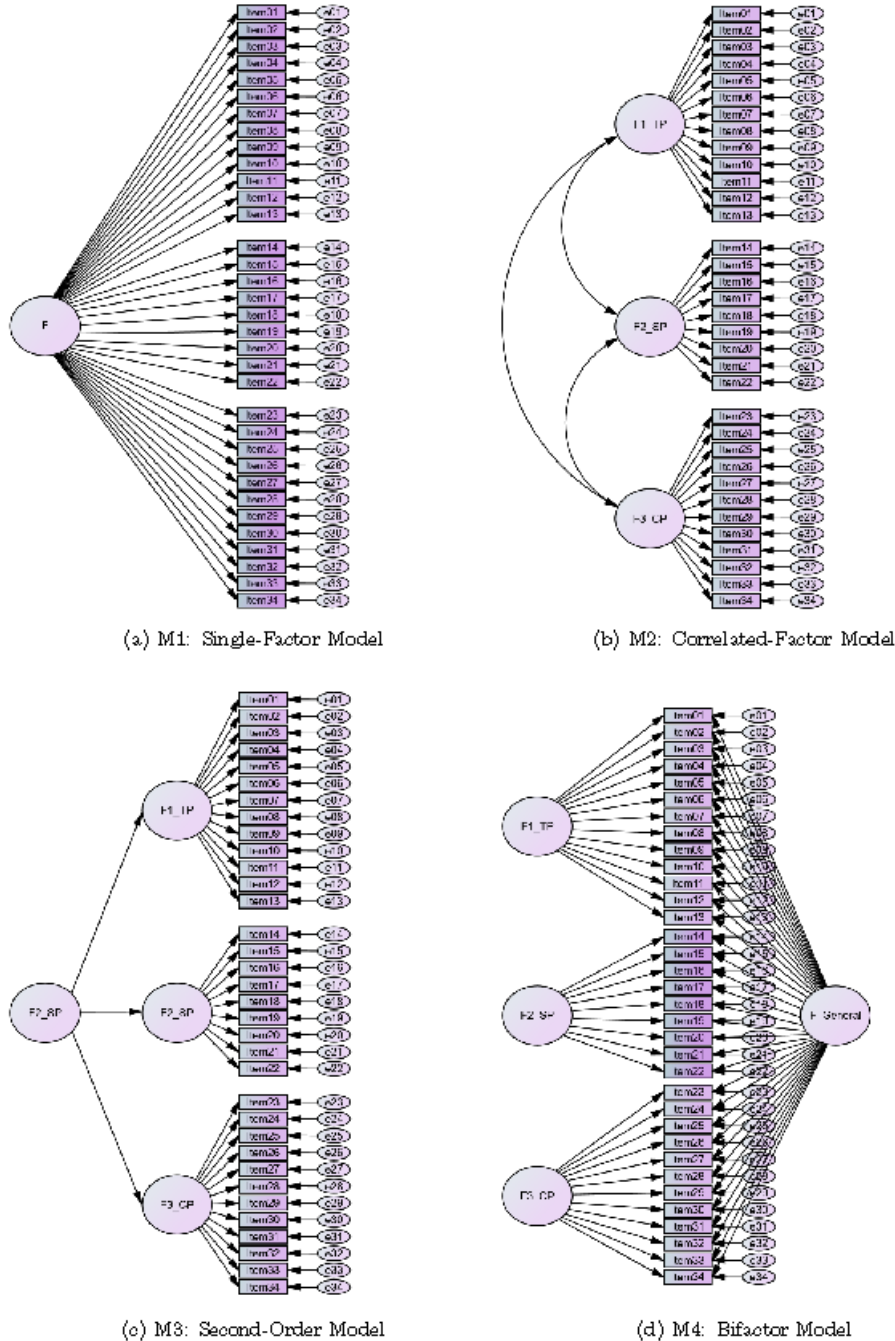
After securing the required Institutional Review Board approval, the study obtained a convenience sample from the participating university in the southeastern US. The sample had a total of 909 graduate students taking online courses in the fall semester of 2014. In January 2015, these 909 students were invited by e-mail to participate in the study through Qualtrics.

To address the common, low response rate issue with online surveys, the study contacted research participants multiple times by e-mail. In the beginning, a massive pre-study notification e-mail was sent to all 909 students, informing them of an upcoming solicitation to participate in a research project on their online learning experiences in Fall, 2014. After the data collection started, additional e-mails followed to remind the participants of responding to the survey. This continued until the data collection came to an end in April 2015.

After addressing the missingness in the responses through listwise deletion, there were 238 participants left who provided complete responses to all 34 CoI items, which led to a student-item ratio of about 7:1, satisfying the criterion that, for stable results, the sample size should be at least six times the number of items (Mundfrom et al., 2005).

Results

Based on the descriptive statistics of the responses, nearly 40% of the CoI items had skewness statistics outside the acceptable range of -1.00 to +1.00, and five items had kurtosis values above the acceptable upper limit of +1.50. It was justified for this study to apply the robust WLS method, instead of treating the categorical responses as if they were continuous (Byrne, 2010; Kline, 2016). The CFA results are found in Tables 2 through 4.

Table 2 presents model fit statistics. All four models demonstrated an adequate fit on CFI, TLI, and AGFI because they were all extremely close to the upper limit of 1.00 for a perfect fit. M4 performed the best on all three fit measures. Regarding SRMR, only M4 and M2 were lower than the threshold of .080 for a good fit. Out of the two models, M4 had the lower SRMR of .051. Regarding RMSEA, M4 had the lowest value of .080 and the lower bound of .074 for its 95% confidence interval was lower than the threshold of .08 for an adequate fit (Byrne, 2010; MacCallum et al., 1996; West et al., 2012). Notably, the thresholds used here for deciding whether a mode fit was adequate have been traditionally designed for the normal-theory maximum likelihood estimation with continuous data. By contrast, this study implemented a WLS estimator with ordinal categorical data. Although there are known methodological issues related to the application of these traditional thresholds to a research context like this, the practice has been widely accepted in the literature and will continue until better alternatives are proposed and established (Xia & Yang, 2019).

Evidently, out of the four models, the bifactor structure showed the best fit as assessed by highest values of CFI, TLI, and AGFI as well as lowest values of SRMR and RMSEA. Next, a Satorra-Bentler scaled χ^2 difference test was run to compare the bifactor structure with each of the other competing structures nested

within the bifactor model. The results indicated the bifactor model was statistically significantly better in fit than each competing structure.

Table 2

Results of Confirmatory Factor Analysis of Four CoI Models

Competing models	X²(df)	CFI	TLI	AGFI	SRMR	RMSEA	RMSEA Lower	RMSEA Upper
Bifactor (M4)	1151.733 (459)	.998	.997	.995	.051	.080	.074	.086
Second-Order factor (M3)	14750.688 (527)	.957	.954	.943	.164	.337	.333	.342
Correlated-factor (M2)	1835.662 (524)	.996	.996	.993	.060	.103	.098	.108
One-Factor (M1)	5844.610 (527)	.984	.983	.977	.116	.206	.202	.211

Table 3 contains the standardized estimates of the bifactor structure. For most items (particularly, CP items), the common variance was explained more by the general factor than by the corresponding subscale domain factor: items 9 (.854 by general vs. .295 by domain), 22 (.825 by general vs. .390 by domain), 23 (.923 by general vs. .115 by domain), and so on. Only several SP items loaded approximately equally on both the general factor and the domain factor: items 17 (.660 by general vs. .644 by domain), 18 (.681 by general vs. .623 by domain), and 20 (.595 by general vs. .630 by domain). Finally, sets of items with high loadings on a subscale may indicate overrepresentation of the content. The fact that items 17, 18, 19, and 20 (all dealing with how comfortable the participant feels about online communication) had high loadings of .644, .623, .712, and .630 on the SP subscale may indicate redundancy from too much content similarity. It is important for a subscale to reflect a conceptually, relatively narrow psychological trait but the construct should not be a mere artifact of asking the same question repeatedly in slightly different ways (Reise, 2012).

Table 3

Bifactor Model Standardized Parameter Estimates for CoI

Item	$\lambda_{General}$	λ_{TP}	λ_{SP}	λ_{CP}	var(e)
01	.779	.569			.069
02	.754	.583			.091
03	.710	.547			.197
04	.662	.568			.239
05	.790	.477			.149
06	.860	.414			.090
07	.826	.364			.186
08	.831	.428			.127
09	.854	.295			.184
10	.816	.341			.218
11	.884	.330			.110
12	.797	.355			.240
13	.673	.443			.351
14	.680		.483		.305
15	.633		.495		.355

16	.602	.449		.436
17	.660	.644		.150
18	.681	.623		.148
19	.616	.712		.114
20	.595	.630		.249
21	.733	.504		.208
22	.825	.390		.167
23	.923		.115	.135
24	.887		.205	.172
25	.864		.177	.222
26	.800		.210	.316
27	.849		.135	.261
28	.908		-.370	.038
29	.880		.054	.223
30	.934		.136	.109
31	.921		.294	.065
32	.882		.412	.053
33	.852		.372	.135
34	.879		.328	.119

Table 4 presents multiple statistics measuring primarily the reliability of the CoI instrument scores under the bifactor model (Reise, 2012; Rodriguez et al., 2016a; 2016b). Among them are model-based reliability measures, measures of construct reliability, and dimensionality.

Table 4

General Factor and Domain Factor Score Reliability Estimates

Statistics	TP (13 items)	SP (9 items)	CP (12 items)	General
α	0.969	0.937	0.968	0.979
ω	0.984	0.966	0.984	0.992
ω_H	0.234	0.388	0.036	0.914
ECV	0.095	0.101	0.029	0.775
H	0.778	0.819	0.478	0.988
Factor Determinacy	0.962	0.966	0.951	0.994
ω, if general factor deleted	0.234	0.387	0.036	0.078
% reduction in reliability, if general factor deleted	76.2	59.9	96.3	92.2

Model-Based Reliability Statistics

Coefficient ω

Coefficient ω for the scale represented the proportion of the variance in the scale total score that was attributable to all sources of common variance (i.e., variance from all common factors: general factor and all domain factors). A high score on the scale ω statistic indicated a highly reliable multidimensional structure that reflected variation on the weighted combination of all common factors. Here, the ω statistic

for the total score was .992, indicating as high as 99.2% of the scale total score variance was due to all common factors.

Coefficient ω for the subscale measured the proportion of the subscale total score variance that was attributable to both the general factor and that domain factor. A high value on the subscale ω statistic indicated a highly reliable multidimensional structure consisting of both the general factor and the domain factor. Here, the subscale $\omega's$ were .984 for TP, .966 for SP, and .984 for CP, suggesting that 96.6% to 98.4% of the subscale total score variance was due to the general factor plus the domain factor.

Coefficient ω_H

Coefficient ω_H for the scale represented the proportion of the scale total score variance that was due to the general factor only. A high score on the scale ω_H statistic indicated the scale total score predominantly reflected the general construct and allowed users to interpret the scale total score as a sufficiently reliable measure of the general factor. Here, ω_H was .914, indicating as high as 91.4% of the scale total score variance was attributed to the general factor after accounting for all domain factors.

Coefficient ω_H for the subscale measured the proportion of the subscale total score variance that was attributed to the domain factor only. A high value on the subscale ω_H statistic indicated the subscale total score predominantly reflected the domain construct and allowed users to interpret the subscale total score as a sufficiently reliable measure of the domain factor. Here, the subscale $\omega_H's$ were .234 (TP), .388 (SP), and .036 (CP), suggesting the subscale total scores mostly did a poor job of reflecting the domain factor and therefore each score was mostly due to the general factor, instead of the domain factor.

Construct Reliability

Construct reliability H represented the proportion of variability in the latent construct explained by its indicator items. A high score on the H statistic indicated the construct was well represented by its indicator items. Here, H = .778 for TP, .819 for SP, .478 for CP, and .988 for the general factor. An H statistic should be at least .70 (or higher) in order for the corresponding latent construct to be adequately represented by its indicators. Therefore, with an H of .988, the general factor was nearly perfectly represented by its 34 indicators and the subscale domains of TP and SP were also represented adequately by their respective indicators. Finally, the domain of CP was not specified reliably given its low H value of .478, indicating the CP domain was not reliably measured by its indicators, and that its results could be unstable and thus not be replicable across studies.

Dimensionality of CoI

The explained common variance (ECV) statistic for the general factor measured the proportion of the common variance explained by both the general and all domain factors which was attributed to the general factor only, and assessed the relative strength of the general factor among all common factors (i.e., essentially measuring the degree of unidimensionality). Here, the general factor ECV statistic was .775, suggesting 77.5% of the common variance across items was explained by the general factor and the remaining 22.5% was spread across the three domain factors. Because the ECV statistic was lower than .85, the instrument was not adequately unidimensional to justify the use of a one-factor model.

Finally, the ECV statistic for an individual item (i.e., IECV) assessed the proportion of item common variance that was explained by the general factor. The closer an IECV was to 1, the stronger the item measured the general construct. If an IECV was greater than .50, the item reflected the general construct more than a domain construct. Here, the IECV statistics ranged from .417 (item 19) to .996 (item 29). The average IECV was .773 with a standard deviation of .161. Two out of the 34 items had an IECV below .50: .417 for item 19 and .477 for item 20, and they measured the domain construct more than they did the general construct. The other 32 items all had an IECV greater than .50 and therefore measured the general construct more than they did the domain construct. Further, eight items had an IECV above .90, indicating they were very strong measures of the general construct.

Discussion

The study conducted a construct-validation of the CoI instrument under CFA followed by further evaluation of the optimal model using primarily model-based reliability measures. The bifactor model identified as being optimal provided a fuller representation of the CoI framework modeling the intersection of all three presences as well as the overlap of each pair of presences. Here, two research questions were proposed and addressed and evidence of construct validity for the instrument was identified.

Regarding the first research question on the fit of each model to the data, the study examined the four competing structures based on commonly used model fit indices. The bifactor structure (M4) was unanimously the optimal one as measured by all five model fit statistics; the other three failed on either one (M2) or two (M1 and M3) of the five criteria. Besides, the CFA results identified items 17, 18, 19, and 20 which may be further examined for content redundancy.

Regarding the second research question on the psychometric properties of the optimal model, the study investigated M4 using primarily model-based reliability measures. Various ω and ω_H statistics provided support for the general factor and demonstrated that the bifactor structure was highly reliable. This finding was echoed by that from various ECV/IECV statistics that showed the general factor played a more important role than did the domain factors in the bifactor model. Finally, the three H statistics for the scale and TP and SP subscales indicated that they were each adequately measured by its indicator items. By contrast, with an H statistic of only .478, the measurement of CP needs more scrutiny.

The finding about CP has both methodological and substantive implications. First, the CP items were probably not measuring cognitive presence effectively. After adjusting for the general factor, the CP factor could hardly continue to exist (Chen et al., 2012). Therefore, the CP items should probably be revised, with the support of subject matter experts, to cover cognitive presence more in-depth. Second, the CP factor scores measuring students' level of cognitive presence should be used carefully. Given high loadings on the general factor but low loadings on the CP factor, it is the general factor scores alone that should be reported (DeMars, 2013) and the domain factor scores could be misleading (Reise et al., 2010; Reise et al., 2007). If policy considerations mandate the reporting of the CP factor scores, users should be reminded that it is the general factor scores that are reliable and meaningful.

The study had limitations which can be grounds for future research. First, the study did not investigate the invariance of the bifactor structure across different groups as specified by common covariates of interest (e.g., gender, course discipline). A future extension could examine if the same bifactor structure continues to hold across those groups (e.g., Dempsey and Zhang, 2019). Second, the study did not test hypotheses on the structural relationships among the common factors of the bifactor model. Another future extension could examine these relationships (e.g., Kozan, 2016). Finally, the study did not assess the predictive validity of the common factors of the bifactor model. Still another future extension may evaluate their predictive validity measured by the associations between the common factors and one or more outside criterion variables such as students' satisfaction with online learning, their academic achievements, and so on (e.g., Rockinson-Szapkiw et al., 2016).

Conclusion

The study conducted a construct revalidation of the CoI instrument for a more refined understanding of its underlying factor structure. The study identified empirical evidence supporting the bifactor model as the optimal structure for providing a reliable and valid representation of the CoI instrument and a fuller representation of the CoI theoretical framework. Therefore, the study recommended the application of the bifactor model to CoI-related research and practice in online education.

References

Arbaugh, B., Cleveland-Innes, M., Diaz, S., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The Internet and Higher Education, 11*(3–4), 133–136. https://doi.org/10.1016/j.iheduc.2008.06.003

Bangert, A. W. (2009). Building a validity argument for the community of inquiry survey instrument. *The Internet and Higher Education, 12*(2), 104–111. https://doi.org/10.1016/j.iheduc.2009.06.001.

Byrne, B. M. (2010). *Structural equation modeling with AMOS* (2nd ed.). Routledge.

Caskurlu, S. (2018). Confirming the subdimensions of teaching, social, and cognitive presences: A construct validity study. *The Internet and Higher Education, 39,* 1–12. https://doi.org/10.1016/j.iheduc.2018.05.002

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing, 13,* 354–378. https://doi.org/10.1080/15305058.2013.799067

Dempsey, P. R., & Zhang, J. (2019). Re-examining the construct validity and causal relationships of teaching, cognitive, and social presence in community of inquiry framework. *Online Learning, 23*(1), 62–79. http://dx.doi.org/10.24059/olj.v23i1.1419

Díaz, S. R., Swan, K., Ice, P., & Kupczynski, L. (2010). Student ratings of the importance of survey items, multiplicative factor analysis, and the validity of the community of inquiry survey. *The Internet and Higher Education, 13*(1–2), 22–30. https://doi.org/10.1016/j.iheduc.2009.11.004

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*(3), 225–241. https://doi.org/10.1177/073428290502300303

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., & Arbaugh, J. B. (2007). Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education, 10*, 157–172. https://doi.org/10.1016/j.iheduc.2007.04.001

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry

framework. *The Internet and Higher Education, 13*(1-2), 31–36.
https://doi.org/10.1016/j.iheduc.2009.10.002

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor
models: Limitations and suggestions. *Intelligence, 62,* 138–147.
https://doi.org/10.1016/j.intell.2017.04.001

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., Čukić, I., de Vries, P., Hatala, M., Dawson, S.,
Siemens, G., & Gašević, D. (2018). Exploring communities of inquiry in massive open online
courses. *Computers & Education, 119,* 44–58. https://doi.org/10.1016/j.compedu.2017.11.010

Kozan, K. (2016). The incremental predictive validity of teaching, cognitive and social presence on
cognitive load. *The Internet and Higher Education, 31,* 11–19.
https://doi.org/10.1016/j.iheduc.2016.05.003

Kozan, K., & Caskurlu, S. (2018). On the nth presence for the community of inquiry framework.
Computers & Education, 122, 104–118. https://doi.org/10.1016/j.compedu.2018.03.010

Kozan, K., & Richardson, J. C. (2014). New exploratory and confirmatory factor analysis insights into the
community of inquiry survey. *The Internet and Higher Education, 23,* 39–47.
https://doi.org/10.1016/j.iheduc.2014.06.002

Ma, Z., Wang, J., Wang, Q., Kong, L., Wu, Y., & Yang, H. (2017). Verifying causal relationships among the
presences of the community of inquiry framework in the Chinese context. *International Review
of Research in Open and Distributed Learning, 18*(6). https://doi.org/10.19173/irrodl.v18i6.3197

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample
size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.
https://doi.org/10.1037/1082-989X.1.2.130

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting
factor analyses. *International Journal of Testing, 5*(2), 159–168.
https://doi.org/10.1207/s15327574ijt0502_4

Olpak, Y. Z., & Kiliç Çakmak, E. (2016). Examining the reliability and validity of a Turkish version of the
community of inquiry survey. *Online Learning, 22*(1), 147–161.
http://dx.doi.org/10.24059/olj.v22i1.990

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research,
47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Rockinson-Szapkiw, A., Wendt, J., Whighting, M., & Nisbet, D. (2016). The predictive relationship among the community of inquiry framework, perceived learning and online, and graduate students' course grades in online synchronous and asynchronous courses. *International Review of Research in Open and Distributed Learning, 17*(3). https://doi.org/10.19173/irrodl.v17i3.2203

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Shea, P., & Bidjerano, T. (2009). Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. *Computers & Education, 52,* 543–553. https://doi.org/10.1016/j.compedu.2008.10.007

Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education, 55*(4), 1721–1731. https://doi.org/10.1016/j.compedu.2010.07.017

Stenbom, S. (2018). A systematic review of the community of inquiry survey. *The Internet and Higher Education, 39,* 22–32. https://doi.org/10.1016/j.iheduc.2018.06.001

Swan, K., & Ice, P. (2010). The community of inquiry framework 10 years later: Introduction to the special issue. *The Internet and Higher Education, 13*(1–2), 1–4. http://dx.doi.org/10.1016/j.iheduc.2009.11.003

Swan, K. P., Richardson, J. C., Ice, P., Garrison, D. R., Cleveland-Innes, M., & Arbaugh, J. B. (2008). Validating a measurement tool of presence in online communities of inquiry. *E-mentor, 2*(24), 1–12. http://www.e-mentor.edu.pl/_xml/wydania/24/543.pdf

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods, 51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2

Yu, T., & Richardson, J. C. (2015). Examining reliability and validity of a Korean version of the community of inquiry instrument using exploratory and confirmatory factor analysis. *The Internet and Higher Education, 25,* 45–52. https://doi.org/10.1016/j.iheduc.2014.12.004

Athabasca University