Measuring Faculty Engagement in Online Formative or Whole-Person Education: A Revised Instrument and Item Response Theory Model

Stanton Wortham
Katrina Borowiec
Deoksoon Kim
Boston College, USA

Abstract
The COVID-19 pandemic reinforced the importance of supporting students' comprehensive well-being when teaching online. One promising approach is formative or whole-person education, which emphasizes wholeness, purpose, and community. We created a scale using a polytomous Item Response Theory modeling approach, measuring the extent to which postsecondary teachers engage in formative education online. To our knowledge, this is the first scale designed to measure this construct. The scale was developed within an exploratory sequential mixed methods study on formative education online that also included semi-structured interviews with 37 faculty members. Results from the qualitative analysis were used to develop initial items. This data-informed process increased the construct validity of the scale. We refined the original item pool through a pilot test using a sample of 308 instructors. This article presents psychometric results for the final, 10-item scale using a sample of 245 instructors. Evaluation of item fit statistics, item trace lines, and the total information curve indicate that the graded response model was appropriate for this scale. The Cronbach's alpha and marginal reliability coefficients for the final scale were .90 and .91, indicating good reliability. Future research can explore how this scale might be adapted for in-person learning environments and other contexts.

Keywords: Formative education, whole-person education, scale development, item response theory

Wortham, S., Borowiec, K., & Kim, D. (2023). Measuring faculty engagement in online formative or whole-person education: A revised instrument and item response theory model. *Online Learning*, *27*(4), 93-121. DOI: https://doi.org/10.24059/olj.v27i4.4033

The COVID-19 pandemic reinforced the importance of supporting students' comprehensive well-being, especially during uncertain times (Wortham et al., 2020). Concerns about students' mental health were well-documented prior to the pandemic (e.g., Oswalt et al., 2015), as were concerns about other aspects of their well-being, such as food insecurity (e.g., Freudenberg et al., 2019). However, the pandemic exacerbated challenges with students' basic needs (Goldrick-Rab et al., 2020), mental health (Elharake et al., 2023; Healthy Minds Network & American College Health Association, 2020), and technology access (Hart et al., 2021; Quezada et al., 2020), while creating new concerns such as how their careers would be impacted by the pandemic (Zhai & Du, 2020).

One comprehensive approach to supporting students' well-being is *formative* or *whole-person education*. Formative education emphasizes three central components: 1) *wholeness*, supporting students' integrated development along intellectual, social-emotional, moral/ethical, and spiritual dimensions; 2) supporting the development of students' sense of *meaning and purpose*; and 3) fostering *community* (Boston College, 2007). Given its focus on human connections, formative education has traditionally been associated with in-person learning environments. Emerging research provides promising evidence that formative education can be done online (Kim et al., 2021).

As defined by Mathes (2020), "Online learning uses the internet as a delivery modality to offer thoughtfully designed, quality, student-focused learning experiences, built on proven best practices that create effective interactions between learners, peers, instructors, and content" (para. 5). Importantly, Mathes's definition centers course design and interactions among teachers, students, and content, while it does not foreground particular tools. We think this focus is important, given the rapidly changing nature of technology. Faculty members play a critical role in ensuring successful online learning by adopting multiple roles, including course facilitator, mentor, manager, designer, and content expert (Martin et al., 2019).

Survey measures are commonly used in research and evaluation to understand individuals' perceptions, behaviors, and experiences. To our knowledge, there are no survey instruments that measure postsecondary teachers' engagement in formative education either in-person or online. The purpose of this study is to develop a scale that measures engagement in formative education online.

The current study is part of a broader exploratory sequential mixed methods study (Creswell, 2015) on formative education online that included qualitative interviews with 37 university faculty skilled in formative education online, which occurred before scale development and pilot testing. This paper presents psychometric findings from a revised survey instrument. We address three research questions:

1.) Is formative education best measured as a unidimensional or multidimensional construct?
2.) Can a graded response Item Response Theory (IRT) model be fit to this scale?
3.) Beyond model fit, what other evidence supports the reliability and validity of the new survey instrument that measures faculty members' engagement in formative education online?

Conceptual Framework and Related Literature

In this section, we provide a foundation for the Concept Map used to develop the Formative Education Online Scale items. Theory and empirical literature about formative education, including our own research and others', were integrated to develop this map. We begin with formative education in general before moving to formative education online.

Formative Education

Many educational philosophies take holistic approaches including those focused on well-being, civic purpose, and character development (Wortham et al., 2020). These philosophies view intellectual development as just one aspect of broader educational aims. Rooted in Jesuit education, "formative" or whole-person education is one example of a comprehensive approach to education, because it facilitates several key aspects of students' development (e.g., intellectual, spiritual, social-emotional, moral/ethical) (Boston College, 2007; O'Malley, 2015). Jesuit educational philosophy claims that caring for students as whole people—sometimes called "cura personalis"—provides a foundation for them to flourish and to offer the best version of themselves in service of others (Geger, 2014; O'Malley, 2015).

Formative education emphasizes three components: 1) wholeness, 2) meaning and purpose, and 3) community (Boston College, 2007; O'Malley, 2015). "Wholeness" describes how formative education promotes integrated student development along intellectual, social-emotional, spiritual, and moral/ethical dimensions (Boston College, 2007). "Meaning and purpose" describes how formative education helps students identify a sense of meaning and purpose in their lives (Boston College, 2007). Finding meaning and purpose involves a process of "discernment" about where one's talents intersect with the needs of humanity (O'Malley, 2015). One crucial step is encouraging students to look beyond extrinsic, instrumental goals, such as status and money. The third component is community. Formative education recognizes that education is not an isolated activity, but rather occurs within a community (Boston College, 2007). The community helps an individual identify larger goals and also provides one crucial end for ethical action. These three components are interrelated. One crucial aspect of holistic development is the discernment of larger life purpose, because a sense of purpose is connected to multiple dimensions of life, informing emotional reactions, underlying fulfilling relationships, and offering a sense of one's place in the larger, moral order, etc. The process of discerning one's purpose and developing holistically is best accomplished within a community.

These three components of formative education can be connected to postsecondary student development theory. Hence we argue that our Formative Education Online Scale should apply to all postsecondary institutions and is not limited to those with Jesuit roots. Kuh (2018), for instance, promotes a holistic educational approach that goes beyond intellectual aspects of student growth to include spiritual, physical, ethical, social, and emotional aspects of student development (i.e., "wholeness"). Chickering and Reisser's (1993) seven vectors of identity development include those that emphasize finding what is meaningful in one's life (i.e., "meaning and purpose") and developing interconnectedness and interpersonal relationships (i.e., "community") (Chickering & Reisser, 1993; Patton et al., 2016). The importance of community—i.e., feeling emotionally connected to and supported by their school and classroom communities—has been well-established (e.g., McMillan & Chavis, 1986). A sense of

community is important to student formation, since formative education is best accomplished through "the help of a companion on the way" (Kolvenbach, 2007, p. 10).

Formative Education Online

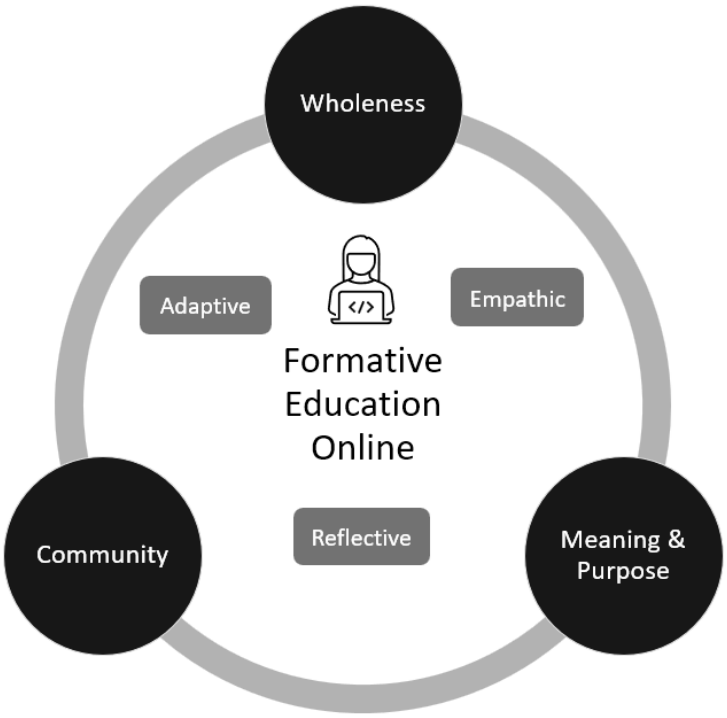
There is limited literature focused on formative education in online settings. However, a recent qualitative study (see Kim et al., 2021) examined how 37 instructors supported students' holistic needs when teaching online. These instructors included faculty across diverse academic fields, including arts and sciences, business, education, law, nursing, social work, theology, and continuing education. The results indicate that faculty members' formative practices can be categorized into three main areas: empathic, reflective, and adaptive. First, faculty continuously demonstrated empathy for their students by reaching out to check in on them, modeling their own vulnerability, and building classroom community. Second, faculty emphasized reflective practices in their classrooms, such as creative activities and practices promoting mindfulness. Third, faculty adapted their instructional approaches to meet students' evolving needs during the pandemic and routinely solicited students' feedback about what was working well and what could use improvement. See Kim et al. (2021) for a detailed account of how these practices directly connect to wholeness, meaning and purpose, and community. For example, *reflective* practices provide opportunities for students to think about broader social issues and how they might address these issues in their future endeavors, providing insight about *meaning and purpose*.

Especially in the context of the COVID-19 pandemic, other scholars have also emphasized the importance of fostering an "ethos of care" (Goin Kono & Taylor, 2021, p. 156) and showing empathy for students' evolving needs (Conklin & Dikkers, 2021; Miller 2021). Researchers have also highlighted the importance of building a sense of community when teaching online (e.g., Borowiec et al., 2021; Castañeda & Selwyn, 2018; Kauffman, 2015; Kilgour et al., 2019; Robinson & Hullinger 2008; Salmon, 2011). In line with the Community of Inquiry Framework (CoI), when teaching online instructors can foster a sense of community through social, cognitive, and teaching presence (Archibald, 2010; Berry, 2019; Garrison et al., 2010; Garrison & Akyol, 2013; Shea & Bidjerano, 2009). By effectively designing learning environments (i.e., teaching presence), instructors can increase students' comfort in the classroom (i.e., social presence) and critical thinking about the course materials (i.e., cognitive presence). Clear, direct one-on-one communication is an important tool for establishing a human connection between students and their instructor in online courses (Berry, 2017; Lowenthal & Dunlap, 2018).

Formative Education Online Concept Map

Figure 1 displays the concept map used to frame our understanding of what it means to engage in formative education online. This framework was used to develop the survey items. As seen in Figure 1, formative education online has three components: 1) wholeness, 2) meaning and purpose, and 3) community. When these three components are integrated, successful formative education online can occur. The three components might be considered the aims of formative education online, which are then accomplished through 1) adaptive, 2) empathic, and 3) reflective teaching practices and philosophies, as shown in our prior empirical research (see Kim et al., 2021).

Figure 1
Concept Map for Formative Education Online

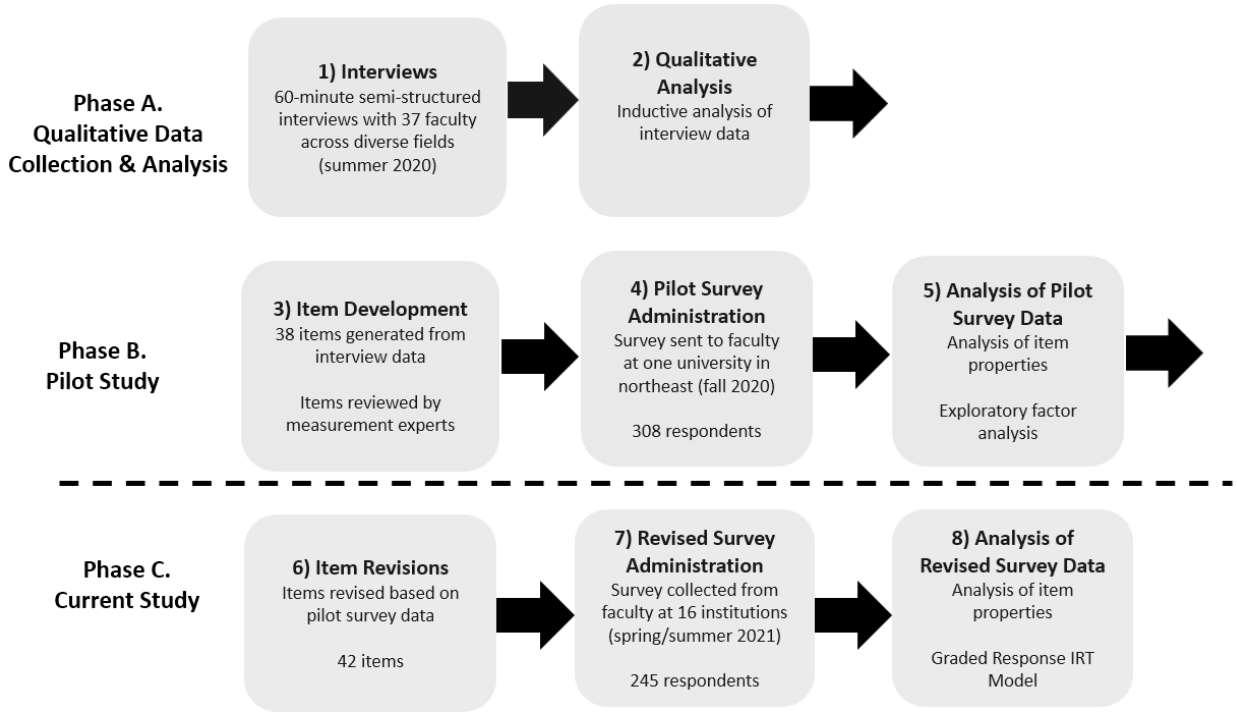


Methodology

Since the Formative Education Online Scale was developed in the context of a broader exploratory sequential mixed methods study (Creswell, 2015), multiple phases of data collection and analysis informed the final 10-item scale. All components of the mixed methods study were approved by the Institutional Review Board.

The instrument was developed through an eight-step process conducted in three phases, A through C, as displayed in Figure 2. Phases A and B provided a foundation for the current study, Phase C. The Methodology section begins with an Overview of the Instrument Development process, before proceeding to discuss the specifics of Data Collection, Participants, and Data Analysis for the final 10-item scale.

Figure 2
Overall Exploratory Sequential Mixed Methods Study Design



Overview of Instrument Development

The current Formative Education Online Scale represents the first measure of formative education online, as far as we know; nor are there any scales measuring in-person engagement in formative education. This creates an exciting opportunity, but it also raises challenges due to the unknown psychometric functioning of the construct. Based on our Concept Map (see Figure 1), our working theory was that engagement in formative education online has three main components: 1) wholeness, 2) meaning and purpose, and 3) community. From a psychometric perspective, it was not immediately clear whether formative education was a unidimensional construct with wholeness, meaning/purpose, and community acting as three aspects of one unified construct, or whether it was a three-dimensional construct with wholeness, meaning and purpose, and community constituting their own dimensions. Nevertheless, we knew it was important for these three aspects of formative education to be represented in the survey items.

In Phase A, Step 1, we interviewed 37 faculty members at one private research university in the United States to understand how faculty provided a formative or whole-person educational approach online (see Kim et al., 2021, for more information about the study). Purposive sampling was used to select these interview participants. Department chairs across the university selected faculty who were exemplary in formative education online, based on their reading of course evaluations and anecdotal information. Formative education is an explicit and widespread part of mission and practice at the university. A qualitative inductive data analysis (Step 2) indicated that faculty members' practices could be organized into three areas: 1) adaptive (i.e.,

faculty were willing to change their courses in response to students' needs), 2) empathic (i.e., faculty recognized the importance of attending to students' social-emotional needs), and 3) reflective (i.e., faculty emphasized teaching practices that encouraged student reflection about course content and the broader world; see Kim et al., 2021, for more information). As discussed in the Conceptual Framework section and depicted in the Concept Map, these themes are also present in related research.

In Phase B, we developed (Step 3), pilot tested (Step 4), and analyzed data from (Step 5) the initial survey instrument. We used the interview data from Phase A to derive item content that reflected the themes of adaptive, empathic, and reflective behaviors that can facilitate wholeness, meaning and purpose, and community. For instance, one item was: "I use material from my class to help students connect how their individual sense of purpose relates to serving other people." This item captures teaching practices that encourage students to reflect on their larger purpose in relation to a community beyond themselves. Items were preceded by the question stem: "When teaching online, how often do you do the following activities?" The 38-item pilot survey was administered online using Qualtrics, and 308 teachers at one university in the northeastern United States completed the survey. No incentive was offered to participate. Items were presented along a five-point Likert scale with the following response options: "never" (1), "rarely" (2), "sometimes" (3), "often" (4), and "very often" (5). Higher scores indicated increased engagement in formative education online.

A Classical Test Theory (CTT) approach was used to analyze the pilot test data (Crocker & Algina, 1986; DeVellis, 2016). The average item difficulty was 3.91, on a scale of 1 to 5, suggesting that the items were "easy" to endorse. Item difficulty was computed as the mean response across all respondents. For the exploratory factor analysis, we explored extraction based on Kaiser's rule to extract factors with eigen values greater than one (DeVellis, 2016), as well as one-factor and three-factor solutions, in accordance with our possible theories about the construct being unidimensional or multidimensional with three categories. Kaiser's rule suggested that eight factors should be extracted, but this solution was ruled out since there was no substantive interpretation of the factors. The one- and three-factor solutions indicated that many items shared a small amount of variance with the other items. We decided to revise the scale before making any decisions about the psychometric structure. Specifically, we wanted to write more "difficult" to endorse survey items. Furthermore, we also wanted to administer the scale to faculty members at diverse institutions.

Phase C represents the current study, which includes item revisions (Step 6), another round of data collection (Step 7), and data analysis using Item Response Theory-based methods (Step 8). Additional details on Steps 6 through 8 are explained below.

Data Collection

Data for the revised survey were collected during spring/summer 2021 from instructors at 16 institutions across the United States, including public community colleges, private four-year colleges, private universities, and public universities. The survey was administered online using Qualtrics. Convenience sampling was utilized. Instructors were invited to participate in the survey via email, and no incentive was offered to participate. A screening question at the beginning of the survey was used to verify that the potential participant had previously taught or

was teaching at least one online, hybrid, or blended course. Most instructors met this qualification because they taught during the COVID-19 pandemic.

The main survey instrument included demographic (e.g., gender, race/ethnicity) and background questions (e.g., years of teaching experience, academic department), the revised 42-item Formative Education Online Scale, and supplemental questions about instructors' experiences teaching online (e.g., benefits, challenges, practices). The supplemental questions are not part of the current study. Several revisions were made to the Formative Education Online Scale between the pilot and the current study. First, five items were dropped and nine items were added. Items were dropped due to lack of clarity. Some items were added to clarify ideas that were poorly captured in the dropped pilot survey items. Based on the pilot test results, we also developed more psychometrically "difficult" items. We added more items related to spiritual and moral practices, since these tended to have lower item means in the pilot test. No changes were made to the question stem or response options.

After the survey administration, we also decided to remove 12 additional items from consideration in the final instrument, yielding 30 items for the data analysis. The project team brainstormed about which items best represented our target construct, in accordance with the Concept Map. A key aim was to remove items that captured more "generic" aspects of good teaching rather than aspects of formative education specifically. From a statistical perspective, we wanted to remove construct-irrelevant variance from the final scale scores by focusing on the most construct-relevant items. In other words, we wanted questions on this measure to be distinct from those that might appear on a general measure of "best teaching" practices. Analogously, a mathematics test that includes reading-heavy tasks may measure both mathematics skills and reading skills. This makes it difficult to know how much of the student's score represents mathematics skills and how much represents reading skills. Our goal was to reduce our measure's item pool and focus on engagement in formative education online, without tapping heavily into related constructs. For example, community is one component of engagement in formative education, but a specific type of community building is most relevant to our construct. Thus, "I strive to create a sense of community in the classroom" was removed, because it reflects general community building. In contrast, the following item was retained: "I encourage students to share their personal life experiences." In this case, the instructor is supporting community building by encouraging openness and vulnerability among students, thus also supporting social-emotional development. Ideally, we would have removed the more "generic" items before the survey administration. But in this case, as in others, scale development was an iterative process.

Participants

A total of 291 instructors responded to the survey. However, to be included in this study, the respondent had to answer at least 15 of the 30 items. 245 instructors met this qualification.

Table 1 presents detailed information about faculty demographic and other characteristics. To summarize: 37% of instructors identified as men, 56% as women, and <0.5% as non-binary. The remaining 7% either preferred not to answer or offered no response. With respect to race/ethnicity, 4% of instructors identified as Asian, 3% as Black or African American, 5% as Hispanic or Latinx, 73% as White, 3% as multiracial, and 1% as some additional race. Eleven percent preferred not to answer or gave no response. When asked

whether they identified as spiritual and/or religious; 32% reported being spiritual; 7% said religious; 24% said both; 31% said neither; and 7% provided no response.

Table 1
Respondent Characteristics

Characteristic	Count	Percent	Characteristic	Count	Percent
Total Participants	245	100%			
Gender			*Academic Discipline*		
Man	91	37%	Arts & Humanities	40	16%
Woman	136	56%	Business	14	6%
Non-binary	1	<.5%	Education	69	28%
I prefer not to answer.	4	2%	Medicine, nursing, other health	34	14%
No response	13	5%	Social sciences	34	14%
			Social work	7	3%
Race/ethnicity			STEM	37	15%
Asian	11	4%	Additional fields	7	3%
Black/African American	7	3%	No response	3	1%
Hispanic/LatinX	12	5%			
White	180	73%	*Teaching Experience*		
Multiracial	7	3%	0-2 Years	18	7%
Additional races	2	1%	3-5 Years	30	12%
Prefer not to answer	12	5%	6-10 Years	39	16%
No response	14	6%	11-15 Years	41	17%
			16-20 Years	33	13%
Spiritual or Religious Identity			21+ Years	84	34%
I am a spiritual person.	78	32%	No response	0	0%
I am a religious person.	17	7%			
Both	58	24%	*Works at Jesuit Institution*		
Neither	76	31%	Yes	75	31%
No response	16	7%	No	170	69%
Tenured or Tenure- Track					
Yes	107	44%			
No	120	49%			
No response	18	7%			

With respect to their academic background, 44% of instructors were on the tenure track, 49% were not, and 7% did not answer the question. Education was the most commonly reported discipline (28%), followed by the arts and humanities (16%), STEM (15%), medicine/nursing/ other health (14%), social sciences (14%), business (6%), and social work (3%). Thirty-one percent of instructors worked at Jesuit institutions.

Instructors were also diverse with respect to their teaching experience: 19% had five years of teaching experience or fewer, while 34% had 21 years or more of teaching experience (see Table 1). Table 2 provides information about the number of completely online and hybrid courses that instructors had taught. All instructors had to have taught at least one course in an online, hybrid, or blended format to participate in the study. Only 5% of instructors never taught a completely online course. The highest proportion of instructors, 27%, reported teaching two-to-three completely online courses. In comparison, the highest proportion of instructors reported teaching zero hybrid courses (47%), although 20% had taught two-to-three.

Table 2

Online and Hybrid Teaching Experience

	Completely Online Courses (Asynchronous or synchronous)		**Hybrid Courses**	
Number of Courses	Count	Percent	Count	Percent
0	12	5%	115	47%
1	33	13%	30	12%
2-3	66	27%	49	20%
4-5	48	20%	18	7%
6-10	50	20%	16	7%
11-15	14	6%	3	1%
16-20	7	3%	4	2%
21 or more	14	6%	7	3%
No response	1	0%	3	1%
Total	245	100%	245	100%

Data Analysis

Data analyses were conducted using IBM SPSS 28 (IBM Corp, 2021) and IRTPRO 6.0 (Vector Psychometric Group, 2022) software. We began the analysis by computing descriptive statistics for each of the 30 items. These included basic frequencies to ensure that all response categories (i.e., "Never" to "Very often") were being utilized by respondents, as well as item difficulties and discriminations using a Classical Test Theory (CTT) approach (Crocker & Algina, 1986; DeVellis, 2016). Using CTT, item difficulty refers to how "difficult" it is for respondents to endorse an item and is computed as the mean response across all respondents. Item discrimination corresponds to how well the item differentiates between respondents of various trait levels and is computed as the corrected item-total correlation—that is, the correlation between the item response and the total test score, excluding that item. From a CTT lens, discrimination below 0.20 is "poor"; between 0.20 and 0.29 is "marginal"; between 0.30 and 0.39 is "reasonably good"; and 0.40 or higher is "very good" (Ebel & Frisbie, 1991, p. 232).

The next series of analyses focused on exploring whether a Graded Response Model (Samejima, 1969) was appropriate for these data. The Graded Response Model (GRM) is part of the family of polytomous Item Response Theory (IRT) models (de Ayala, 2009). Polytomous IRT models are statistical models, also known as latent trait models, that use ordinal regression

to predict a respondent's likelihood of endorsing an item at a certain level (e.g., "Never," "Rarely"), given their overall trait level. For the Formative Education Online Scale, the latent trait is "engagement in formative education online."

The GRM can be represented statistically as follows (de Ayala, 2009, p. 219):

$$P^*_{x_j}(\theta) = \frac{e^{\alpha_j(\theta - \delta_{x_j})}}{1 + e^{\alpha_j(\theta - \delta_{x_j})}}$$

A separate GRM is computed for each item in a scale. This formula represents an individual's probability, *P**, of responding to a specific item, *j*, at a response level of x_j or higher, given their estimated trait level, θ. The parameter, α_j, represents the item discrimination parameter, which indicates how well the item differentiates between respondents of different trait levels (i.e., higher α values indicate improved discrimination). Finally, δ_{x_j} is the boundary or threshold between two adjacent response categories, *x* and *x*-1. For the Formative Education Online Scale, an example would be the boundary between "Sometimes" and "Rarely." Additionally, δ_{x_j} represents the amount of engagement in formative education online (i.e., the trait level) at which a respondent has a 50% probability of responding at that category, *x*, or higher for item *j* (de Ayala, 2009).

IRT models also have three major assumptions (de Ayala, 2009; Hambleton et al., 1991). The first assumption is unidimensionality, which means that there is one underlying trait. This assumption will be examined using principal axis factoring in SPSS. There should be one underlying factor or one dominant factor that explains a large proportion of variance (Hambleton et al., 1991). Some level of violation is often observed in practice (de Ayala, 2009). If the construct is multidimensional, then a multidimensional IRT model must be used or a separate GRM can be computed for each dimension. The second assumption is conditional independence, which means that a person's response to each item is not dependent on their responses to other items in the scale. This assumption will be examined using Local Dependency Chi-square statistics provided by IRTPRO. Values above 10 indicate that there may be issues with local dependency that deserve further examination (Vector Psychometric Group, 2020). The third assumption is that there is alignment between the data and the GRM model, which will be examined using overall model fit statistics, item fit statistics, and examination of the item trace lines. Overall model fit will be examined using the Root Mean Square Error of Approximation (RMSEA) statistic. Hu and Bentler (1999) suggest that an RMSEA statistic of .06 or below indicates "good" fit. Item fit will be examined using chi-square item fit diagnostics. Typically, an alpha level of .05 is used to detect some degree of misfit; however, if the p-value is above .01, then the misfit can be considered negligible (Vector Psychometric Group, 2020).

Finally, the overall reliability and validity of the final scale will be examined. Reliability will be examined using both Cronbach's alpha as a measure of internal consistency (DeVellis, 2016) and marginal reliability, which indicates the average reliability of the scale across the latent trait (de Ayala, 2009). Cronbach's alpha of 0.7 or higher is considered acceptable (Nunnally, 1978). We will use a comparable standard to evaluate marginal reliability. The construct validity of the scale is supported through its rigorous development process. The Concept Map was developed using both theory and empirical research. An early version of the

items was reviewed by psychometricians and their feedback was integrated into the survey. Finally, we will test the "known groups" validity (DeVellis, 2016) of the scale by comparing mean IRT-based total scores for instructors at Jesuit institutions to those at all other institutions using an independent samples t-test. We anticipate that this scale will have utility at all postsecondary institutions. However, we expect that instructors at Jesuit institutions will score higher, on average, given that the scale was developed based on a Jesuit approach to formative education online. Hence, instructors at Jesuit institutions are expected to engage in these practices more often, on average, as part of their institutional cultures.

Results

The results are organized into two main sections. The first section reviews the initial analysis of the revised scale, and the second section provides descriptive statistics, reliability, and validity evidence for the final scale.

Initial Analysis of 30 Items in the Revised Scale

Table 3 provides descriptive statistics for the 30 items in the revised scale, including item difficulties (i.e., means) and discriminations (i.e., corrected item-total correlations) using a CTT approach. Item difficulties ranged from 2.70 to 4.51, with an average of 3.80. The average item difficulty for the revised scale was slightly lower than for the pilot (3.80 vs. 3.91). Item discrimination ranged from 0.44 to 0.74, with an average of 0.58. All items had "very good" discrimination (Ebel & Frisbie, 1991).

Frequency statistics for the 30 items resulted in two items, formed26 and formed36, being removed from consideration for the final scale. The first item was removed because no respondent utilized the "never" category. A second item (formed36) was removed because only one respondent utilized the "never" and "rarely" categories, respectively. We opted to remove these items because the response categories did not seem appropriate for the items, given the low utilization of certain categories.

Table 3

Descriptive Statistics for 30 Items in Revised Scale

Item	N	Min.	Max.	Mean	SD	Corrected Item-Total Correlation
formed01	244	1	5	4.12	0.92	0.58
formed03	245	1	5	3.80	1.04	0.58
formed04	245	1	5	3.13	1.17	0.57
formed06	241	1	5	3.70	1.09	0.64
formed07	244	1	5	2.70	1.27	0.57
formed08	244	1	5	3.88	0.99	0.61
formed10	244	1	5	2.92	1.26	0.59
formed11	244	1	5	3.91	1.04	0.50
formed12	245	1	5	3.96	1.04	0.67

formed13	245	1	5	3.83	1.11	0.52
formed15	243	1	5	3.57	1.21	0.74
formed17	245	1	5	4.04	1.01	0.46
formed18	245	1	5	3.20	1.23	0.49
formed20	245	1	5	3.58	1.09	0.74
formed21	245	1	5	4.25	0.87	0.58
formed22	243	1	5	4.06	1.18	0.48
formed23	244	1	5	4.11	0.98	0.57
formed24	243	1	5	4.15	0.89	0.58
formed26	244	2	5	4.05	0.84	0.59
formed32	240	1	5	3.60	0.97	0.44
formed33	240	1	5	4.30	0.90	0.63
formed34	240	1	5	3.90	1.22	0.51
formed35	239	1	5	3.72	1.13	0.45
formed36	240	1	5	4.38	0.67	0.52
formed37	239	1	5	3.93	1.02	0.68
formed38	238	1	5	3.87	1.01	0.60
formed39	237	1	5	3.73	1.16	0.68
formed40	236	1	5	4.51	0.81	0.50
formed41	235	1	5	4.20	1.01	0.55
formed42	234	1	5	2.87	1.35	0.66

Table 4

Principal Axis Factoring Results for 28-Items Without Rotation

Item	**Factor**					
	1	2	3	4	5	6
formed15	0.77					
formed20	0.76					
formed39	0.70					
formed37	0.70					
formed12	0.69					
formed42	0.69		-0.36			
formed06	0.66					
formed33	0.65					
formed08	0.64					
formed38	0.63					
formed10	0.62					
formed23	0.60	0.50	0.40			
formed04	0.60	-0.37				
formed03	0.60					

formed07	0.60			-0.32		
formed24	0.60	0.41	0.33			
formed21	0.59					
formed01	0.59					
formed41	0.58					
formed13	0.56	-0.56	0.35			
formed11	0.53					
formed34	0.52				-0.46	
formed40	0.52			0.36		
formed22	0.50	0.35	0.33			
formed18	0.50					
formed35	0.48	-0.36	0.34			
formed17	0.47					
formed32	0.45					

Note. Only factor loadings of .3 or higher are displayed.

Table 4 displays the principal axis factoring results. Six factors were extracted using Kaiser's rule, to extract based on the number of eigen values greater than 1 (DeVellis, 2016). The findings support the idea of a dominant factor, since all 28 items loaded on the first factor with a loading of 0.45 or higher. Moreover, the first factor explained 37% of the shared variance, while the other five each explained 5% or less variance. Notably, the fourth, fifth, and sixth factors were weak, explaining 3% or less variance each. Zero items loaded on the sixth factor. A solution with more than three factors did not seem appropriate. A forced three-factor solution with a Promax rotation indicated that all three factors are moderately to strongly correlated with one another (factors 1 and 2: r=.67; factors 1 and 3: r=.61; and factors 2 and 3: r=.49). This provides further support for a dominant first factor. Any violations of the unidimensionality assumption of IRT seem modest. As de Ayala (2009) noted, there is typically some level of violation of this assumption in practice.

Based on these results, we proceeded with an assumption of unidimensionality in the formative education online construct. It was then appropriate to proceed to IRT-based modeling. After further consideration, the idea that the construct was unidimensional also aligned better with theory. Although engagement in formative education online contained three key components (wholeness, meaning/purpose, community), these components best capture the nature of formative education online when working harmoniously together.

Through an iterative process, we removed 18 additional survey items and this yielded a final scale of 10 items. Since the construct was presumed unidimensional at this point in the development process, we could reduce the number of total items since we only needed to compute one score. It was also important to reduce respondents' burden, so that the scale can more easily be utilized. When selecting the final 10 items, we balanced a variety of considerations. We first reduced the item pool to 16 items based on the aforementioned psychometric analyses (e.g., item difficulties and discriminations, factor analysis results) and the item content, so that selected items would still capture wholeness, meaning and purpose, and

community while embodying adaptive, empathic, and reflective teaching practices. After selecting these 16 items, we ran an initial GRM in IRTPRO 6.0 and plotted trace lines. We then selected the final 10 items based on their content and the trace lines. The ideal trace lines were those in which each category clearly had a range of theta where a respondent was most likely to select that category, suggesting that the categories were being utilized by respondents as intended.

Table 5

Overview of Item Revision Process

Study Phase and Specific Survey Task	Number of Items and Explanation for Reduction
Pilot Survey Administration	**38 items**
Revised Survey Administration	**42 items** Of the original 38 items, five items were dropped due to lack of clarity. Nine items were added to capture missing aspects of the construct and to add more psychometrically "difficult" items.
Revised Survey Post-Administration	**30 items** Of the 42 items in the Revised Survey, the research team removed 12 items that seemed too "generic" and potentially captured construct-irrelevant variance.
Revised Survey Data Analysis: Item Difficulties, Discriminations, and Frequencies	**28 items** Of the 30 items, two were removed because respondents were not using all response options, suggesting that the response categories were not appropriate for the items.
Revised Survey Data Analysis: Initial Item Reduction Phase	**16 items** Through an iterative process, we reduced the 28 items to 16. After deciding the construct was unidimensional, 28 items were not needed. It was important to reduce the number of items in order to reduce respondents' burden. Item reduction was done with reference to a combination of item content, item difficulty, item discrimination, and factor analysis results.
Revised Survey Data Analysis: Additional Item Reduction Phase and IRT Model	**10 items** We selected the final 10 items with reference to their content combined with the item trace lines produced through IRT modeling. The ideal trace lines were those in which each response category had a range of theta where a respondent was most likely to select that response.

Final 10-Item Scale

In this section, we provide information about the reliability and validity of the final 10-item Formative Education Online Scale. The scoring procedures for the final 10-item scale can be found in the Appendix. The final score is scaled on the T-Score metric with a mean of 50 and a standard deviation of 10.

The final 10-items are presented in Table 6, along with their item difficulties (i.e., item means) and discriminations (i.e., corrected item-total correlations) using a CTT-based calculation. The item difficulties ranged from 2.70 to 3.96, with an average difficulty of 3.41. As desired, this represents a more "difficult" scale than we had after the pilot test. Item discrimination ranged from 0.59 to 0.76, with an average of 0.66. These discrimination values are all considerably higher than the 0.40 criterion for "very good" discrimination suggested by Ebel and Frisbie (1991).

Table 6

Descriptive Statistics for Items in Final Scale

Item Description	Item	N	Mean	SD	Corrected Item-Total Correlation
I incorporate assignments that ask students to consider moral dilemmas.	formed04	245	3.13	1.17	0.59
I encourage students to share their personal life experiences.	formed06	241	3.70	1.09	0.63
I welcome students to contribute their spiritual or religious beliefs and values in classroom discussions.	formed07	244	2.70	1.27	0.63
I provide opportunities in class for mindfulness and/or contemplation.	formed10	244	2.92	1.26	0.63
I encourage students to incorporate what matters to them in the course assignments.	formed12	245	3.96	1.04	0.63
I use material from my class to help students connect how their individual sense of purpose relates to serving other people.	formed15	243	3.57	1.21	0.76
I design assignments and facilitate conversations that help students identify what is meaningful in their lives.	formed20	245	3.58	1.09	0.72
I structure my courses to encourage students to think beyond their personal experiences and toward their ultimate contribution to a greater good.	formed37	239	3.93	1.02	0.64
I encourage students to develop their own moral compass.	formed39	237	3.73	1.16	0.66
My course attends to the integration of mind and spirit as a component of self-discovery.	formed42	234	2.87	1.35	0.71

A graded response model was then computed for these 10 items using IRTPRO 6.0. Table 7 displays the factor loadings for each item in the final scale. The factor loadings were high, ranging from 0.64 to 0.85.

Table 7
Factor Loadings for Final 10-Item Scale

Item	λ_1	*s.e.*
formed04	0.64	0.08
formed06	0.70	0.08
formed07	0.68	0.08
formed10	0.70	0.08
formed12	0.74	0.08
formed15	0.85	0.06
formed20	0.84	0.06
formed37	0.75	0.08
formed39	0.74	0.08
formed42	0.81	0.06

The RMSEA for the overall model is 0.06, indicating good overall fit (Hu & Bentler, 1999). Table 8 displays item level fit statistics. Only one item, formed04, had a p-value below .05, which indicates misfit. However, p-values above .01 indicate that the misfit can be considered negligible (Vector Psychometric Group, 2020).

Table 8
S-X^2 Item Level Diagnostic Statistics

Item	X^2	*d.f.*	*p*
formed04	91.78	68	0.029
formed06	56.04	56	0.475
formed07	67.56	62	0.293
formed10	73.18	63	0.178
formed12	59.89	51	0.184
formed15	49.16	48	0.428
formed20	50.86	45	0.253
formed37	52.91	48	0.290
formed39	69.05	55	0.096
formed42	62.79	56	0.248

Table 9 provides a standardized local dependency chi-square statistic, LD χ^2, for each item pair. Values above 10 indicate that there may be issues with local dependency that deserve further examination (Vector Psychometric Group, 2020). The highest value for these data is 3.5, indicating that local dependency is not a problem with this scale.

Table 9
Standardized LD X^2 Statistics

Item	1	2	3	4	5	6	7	8	9
formed04									
formed06	1.5								
formed07	-0.7	-0.8							
formed10	0.3	-0.0	-0.0						
formed12	-0.8	0.0	-0.9	3.5					
formed15	-1.2	-0.8	3.0	0.9	0.9				
formed20	0.3	-1.6	0.8	2.3	1.9	2.4			
formed37	-0.9	-0.2	1.9	-0.1	-0.9	-0.4	-0.7		
formed39	-0.3	-1.2	1.8	0.8	0.2	0.6	0.9	1.9	
formed42	-0.3	-0.7	2.1	0.6	3.7	1.0	0.8	0.4	2.2

Table 10 provides information about the item parameters: discrimination, a item thresholds, b_1 to b_4, and item intercepts, c_1 to c_4. This information can be used to predict an individual's likelihood of providing a particular response (e.g., "Never"), given their latent trait level (i.e., amount of engagement in formative education online). Item formed15 has the highest discrimination, meaning that answering this item provides the most information about an instructor's amount of engagement in formative education online.

Table 10

Graded Model Item Parameter Estimates for Final 10-Item Scale, logit: $a(\theta - b)$ and logit: $a\theta + c$

	Item Discrimination [SE]	***Item Thresholds [SE]***				***Item Intercepts [SE]***			
Item	a	b_1	b_2	b_3	b_4	c_1	c_2	c_3	c_4
formed04	1.41 [0.18]	-2.10 [0.25]	-0.82 [0.14]	0.46 [0.17]	1.65 [0.26]	2.97 [0.30]	1.15 [0.22]	-0.66 [0.21]	-2.34 [0.26]
formed06	1.67 [0.22]	-2.82 [0.33]	-1.54 [0.17]	-0.28 [0.13]	0.79 [0.18]	4.71 [0.50]	2.57 [0.30]	0.47 [0.23]	-1.31 [0.24]
formed07	1.57 [0.21]	-1.22 [0.15]	-0.14 [0.13]	0.90 [0.19]	1.81 [0.28]	1.92 [0.26]	0.22 [0.21]	-1.42 [0.23]	-2.84 [0.29]
formed10	1.67 [0.22]	-1.52 [0.17]	-0.43 [0.12]	0.58 [0.16]	1.51 [0.24]	2.53 [0.29]	0.71 [0.23]	-0.97 [0.23]	-2.53 [0.28]
formed12	1.86 [0.25]	-2.78 [0.33]	-1.66 [0.17]	-0.75 [0.12]	0.48 [0.15]	5.16 [0.56]	3.09 [0.35]	1.39 [0.27]	-0.89 [0.24]
formed15	2.77 [0.39]	-1.74 [0.15]	-1.03 [0.10]	-0.23 [0.10]	0.71 [0.16]	4.81 [0.58]	2.87 [0.44]	0.64 [0.34]	-1.96 [0.32]
formed20	2.58 [0.35]	-2.06 [0.20]	-1.16 [0.11]	-0.17 [0.10]	0.92 [0.17]	5.33 [0.59]	2.99 [0.42]	0.43 [0.30]	-2.38 [0.32]
formed37	1.90 [0.27]	-2.73 [0.33]	-1.76 [0.18]	-0.74 [0.11]	0.55 [0.16]	5.19 [0.58]	3.35 [0.38]	1.41 [0.27]	-1.05 [0.23]
formed39	1.88 [0.26]	-2.24 [0.25]	-1.34 [0.15]	-0.35 [0.11]	0.68 [0.16]	4.21 [0.45]	2.52 [0.32]	0.66 [0.24]	-1.27 [0.23]
formed42	2.37 [0.31]	-1.05 [0.12]	-0.27 [0.10]	0.59 [0.14]	1.22 [0.19]	2.48 [0.34]	0.63 [0.28]	-1.39 [0.28]	-2.88 [0.33]

Note. The standard error [SE] for each parameter is presented in brackets. The subscript 1 corresponds to "Rarely," 2 to "Sometimes," 3 to "Often," and 4 to "Very Often."

The trace lines for all 10 items are presented in Figures 3a through 3b. The categories of 0, 1, 2, 3, and 4 correspond to the response options "Never," "Rarely," "Sometimes," "Often," and "Very Often," respectively. For example, the leftmost chart in Figure 3a corresponds to item "formed04." The most probable response to this item for a postsecondary teacher with a theta (i.e., amount of engagement of formative education online) of around -2.0 or below is "Never." In comparison, postsecondary teachers with a theta level around 1.5 or higher are most likely to respond "Very Often" on this item.

Figure 3a

Trace Lines for Items: Formed04, Formed06, Formed07, and Formed10 (Respectively)

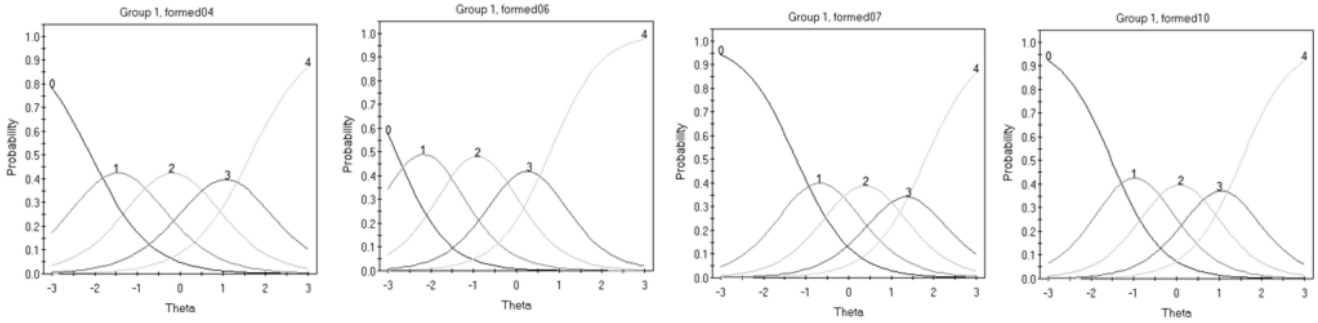


Figure 3b

Trace Lines for Items: Formed12, Formed15, Formed20, and Formed37 (Respectively)

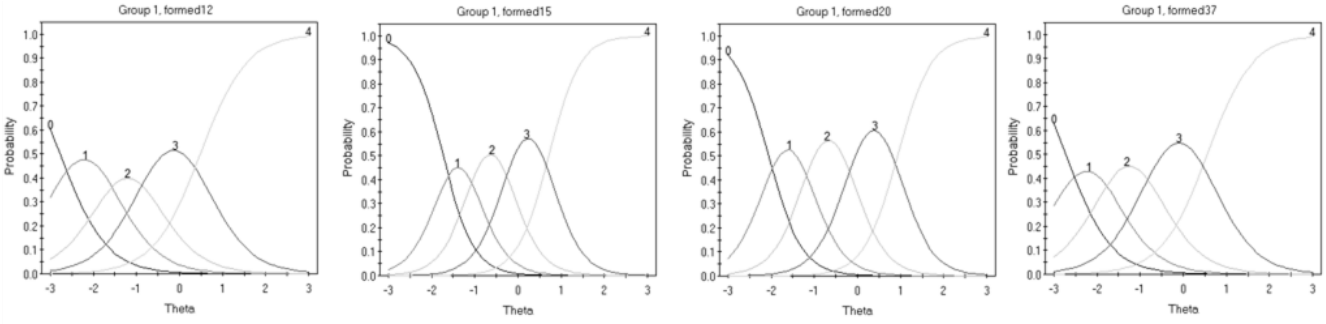
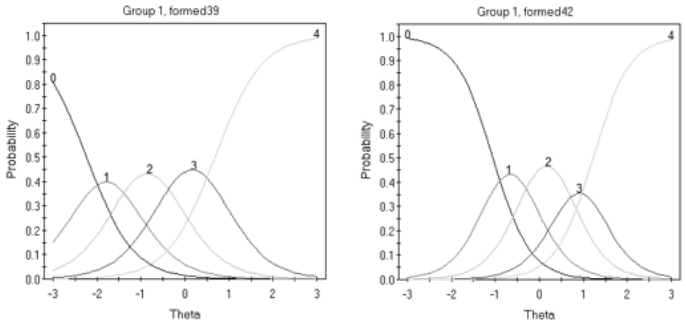


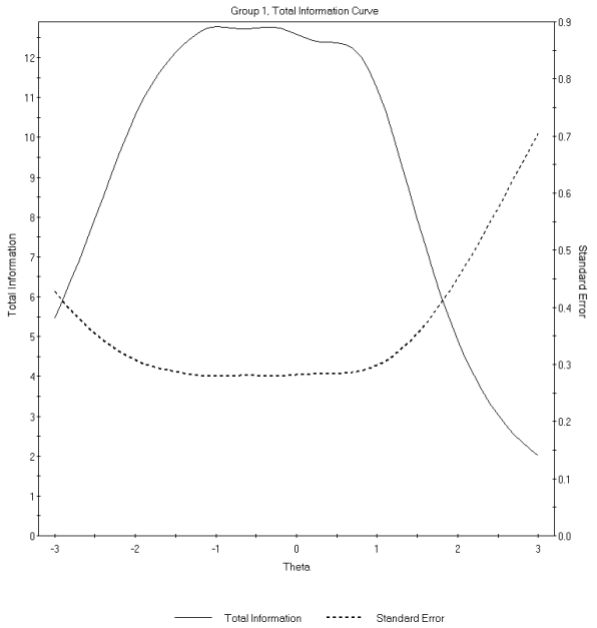
Figure 3c

Trace Lines for Items: Formed39 and Formed42 (Respectively)



Cronbach's alpha reliability and marginal reliability were both high at .90 and .91, respectively, indicating very good overall reliability for the scale. Figure 4 displays the Total Information Curve for the Formative Education Online Scale. From this graph, we can see that the scale provides a high amount of information relative to the amount of error. The only range of the trait where more information might be needed is at the upper range of the scale. In other words, more psychometrically "difficult" items are needed to better capture very high trait levels.

Figure 4
Total Information Curve for the Formative Education Online Scale



The known-groups validity was supported. The independent samples t-test indicated that instructors from Jesuit institutions (N=75, M=54.55, SD=8.91) had significantly higher scores on average than those from non-Jesuit institutions (N=170, M=47.97, SD=9.03) (t=5.27, df=243, p<.001). On average, instructors at Jesuit institutions scored 6.58 points higher on the formative education online scale than instructors at other institutions (Cohen's d=0.73).

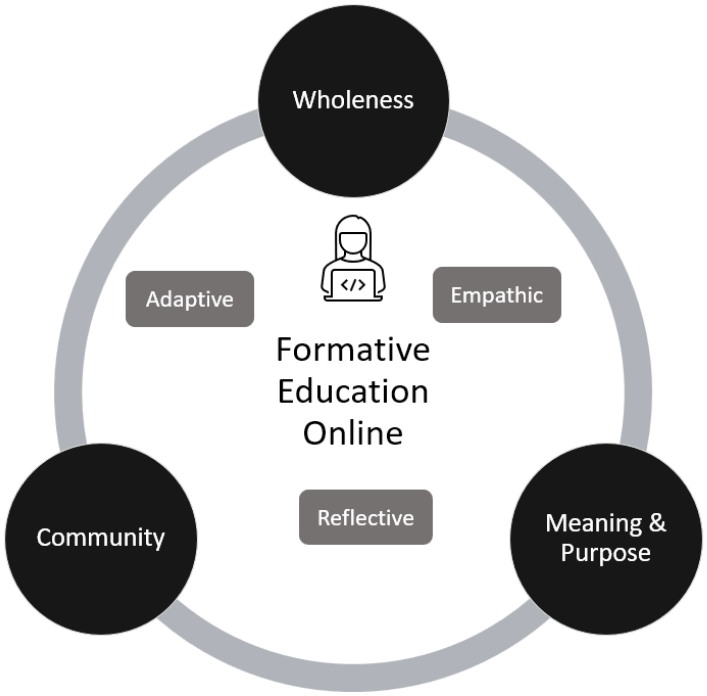
Discussion

The COVID-19 pandemic brought considerable challenges to educators, but it also created opportunities to embrace new teaching practices. The pandemic also made clear the importance of attending to students' holistic needs when teaching online (Borowiec et al., 2021; Conklin & Dikkers, 2021; Goin Kono & Taylor, 2021; Kim et al., 2021; Miller 2021). Formative or whole-person education provides one promising framework for supporting students' well-being online (Wortham et al., 2020).

Colleges should evaluate the extent to which faculty members engage in these formative practices when teaching online. A survey is perhaps the most straightforward and cost-effective approach to evaluating such practices. To our knowledge, our work represents the first scale designed to measure the extent to which faculty members engage in formative education online. Results indicate that a psychometrically sound measure of engagement in formative education online can be developed with strong psychometric properties. Specifically, the Graded Response Model (Samejima, 1969) was fit to the data, which is an IRT-based model. Our empirical results, together with our theoretical understanding of the construct, suggest that engagement in

formative education online is unidimensional. Wholeness, meaning/purpose, and community (Boston College, 2007) represent specific aspects of one unified construct.

Figure 5
Formative Education Online Scale Concept Map and Items



Formative Education Online Scale

1. I incorporate assignments that ask students to consider moral dilemmas.
2. I encourage students to share their personal life experiences.
3. I welcome students to contribute their spiritual or religious beliefs and values in classroom discussions.
4. I provide opportunities in class for mindfulness and/or contemplation.
5. I encourage students to incorporate what matters to them in the course assignments.
6. I use material from my class to help students connect how their individual sense of purpose relates to serving other people.
7. I design assignments and facilitate conversations that help students identify what is meaningful in their lives.
8. I structure my courses to encourage students to think beyond their personal experiences and toward their ultimate contribution to a greater good.
9. I encourage students to develop their own moral compass.
10. My course attends to the integration of mind and spirit as a component of self-discovery.

The scale had high reliability, as measured via both Cronbach's alpha and marginal reliability. Moreover, the Total Information Curve indicated that the scale provided a high level of information across a broad range of the trait. The construct validity of this measure is supported through the development process, in which both theory and empirical data were used to develop a Concept Map that was the foundation for item development. Figure 5 displays the Concept Map and the final 10 items to the right. A review of the item content in the final scale shows strong alignment between the Concept Map and the scale. Some elements may be more explicit, but all are represented. For example, "I encourage students to incorporate what matters to them in the course assignments" reflects meaning and purpose, but also adaptive practices. Instructors are incorporating pedagogy that is flexible. This flexibility allows students to explore their interests with one goal development of a sense of meaning and purpose. Known groups validity was also supported in that instructors at Jesuit institutions had higher engagement in formative education online, on average, than instructors at other institutions. While this scale may have particular relevance to Jesuit institutions, our conceptual framework and review of the related research indicates that the concepts represented in the Concept Map are supported in the general postsecondary research literature and theory (e.g., Castañeda & Selwyn, 2018; Chickering & Reisser, 1993; Conklin & Dikkers, 2021; Goin Kono & Taylor, 2021; Kauffman, 2015; Kilgour et al., 2019; Kuh, 2018; Miller, 2021; Patton et al., 2016; Robinson & Hullinger 2008; Salmon, 2011).

While the scale has many strengths, two limitations should be noted. First, although robust enough to support our analyses, the sample size is relatively small for IRT-based psychometrics (*N*=245). Nevertheless, there were no issues with model convergence. It may be helpful to replicate these findings using a larger sample. Second, as with any self-report measure, these data rely on the perceptions of the respondents.

Conclusion

The current study provides promising evidence for the reliability and validity of our first-of-its-kind Formative Education Online Scale. Future research can explore how this scale might be adapted for in-person learning environments and other contexts, such as K-12 education levels or in educational institutions outside the United States. Additional data can be collected to explore how this scale correlates with other measures that are of interest to educators, researchers, and policymakers—e.g., those measuring students' sense of belonging in the classroom, student course outcomes, and persistence to graduation.

Educators, policymakers, parents, and students increasingly recognize the importance of a whole-person approach to education. Young people are simultaneously undergoing consequential development along various dimensions—intellectual, emotional, relational, ethical, spiritual, etc. If educators ignore this and focus only on content knowledge and vocational skills, young people often suffer. Recognizing this, countries around the world are increasingly attending to student well-being and whole-person development. The rapid growth in online learning, spurred in part by the pandemic, complicates these recent efforts, however. We now have evidence that whole person education can be done in online environments, but in order to do so effectively educators and policymakers need better data. Our Formative Education Online Scale can provide one useful tool in this important effort.

Acknowledgements
The authors acknowledge Dr. Larry Ludlow and Dr. Henry Braun for their early feedback on the survey as well as all research participants.

Declarations
This study was approved by the Boston College Institutional Review Board.
The authors declare no conflicts of interest.

References

Archibald, D. (2010). Fostering the development of cognitive presence: Initial findings using the community of inquiry survey instrument. *Internet & Higher Education, 13*(1/2), 73–74. https://doi.org/10.1016/j.iheduc.2009.10.001

Berry, S. (2017). Building community in online doctoral classrooms: Instructor practices that support community. *Online Learning, 21*(2), n2. http://dx.doi.org/10.24059/olj.v21i2.875

Berry, S. (2019). Teaching to connect: Community-building strategies for the virtual classroom. *Online Learning, 23*(1), 164-183. http://dx.doi.org/10.24059/olj.v23i1.1425

Borowiec, K., Kim, D., Wang, L., Kim, J., & Wortham, S. (2021). Supporting holistic student development through online community building. *Online Learning, 25*(4), 125-155. https://olj.onlinelearningconsortium.org/index.php/olj/article/view/2882

Boston College. (2007). *The journey into adulthood: Understanding student formation*. http://www.bc.edu/content/dam/files/offices/mission/pdf1/umm1.pdf

Castañeda, L., & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education, 15*(1), 22. https://doi.org/10.1186/s41239-018-0109-y

Chickering, A. W., & Reisser, L. (1993). *Education and identity* (2nd ed.). Jossey-Bass Inc.

Conklin, S. & Dikkers, A. G. (2021). Instructor social presence and connectedness in a quick shift from face-to-face to online instruction. *Online Learning, 25*(1), 135-150. https://doi.org/10.24059/olj.v25i1.2482

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Sage Publications.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth Publishing.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Publications.

DeVellis, R. F. (2016). *Scale development: Theory and applications*. Sage Publications.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed). Prentice Hall.

Elharake, J. A., Akbar, F., Malik, A. A., Gilliam, W., & Omer, S. B. (2023). Mental health impact of COVID-19 among children and college students: A systematic review. *Child Psychiatry & Human Development, 54*, 913-925. https://doi.org/10.1007/s10578-021-01297-1

Freudenberg, N., Goldrick-Rab, S., & Poppendieck, J. (2019). College students and SNAP: The new face of food insecurity in the United States. *American Journal of Public Health, 109*(12), 1652-1658. https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2019.305332

Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed.) (pp. 104-119). Routledge.

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education, 13*(1), 5–9. https://doi.org/10.1016/j.iheduc.2009.10.003

Geger, F. B. T. (2014). Cura personalis: Some Ignatian inspirations. *Jesuit Higher Education, 3*(2), 6–20.

Goin Kono, K., & Taylor, S. (2021). Using an ethos of care to bridge the digital divide: Exploring faculty narratives during a global pandemic. *Online Learning, 25*(1), 151-165. https://doi.org/10.24059/olj.v25i1.2484

Goldrick-Rab, S., Coca, V., Kienzl, G., Welton, C. R., Dahl, S., & Magnelia, S. (2020). *#RealCollege during the pandemic: New evidence on basic needs insecurity and student well-being*. The Hope Center.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Hart, C.M.D., Xu, D., Hill, M., & Alonso, E. (2021). COVID-19 and community college instructional responses. *Online Learning, 25*(1), 41-69.

Healthy Minds Network, & American College Health Association. (2020). *The impact of COVID-19 on college student well-being*. https://healthymindsnetwork.org/wp-content/uploads/2020/07/Healthy_Minds_NCHA_COVID_Survey_Report_FINAL.pdf

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

IBM Corp. (2021). *IBM SPSS Statistics for Windows* (Version 28.0) [Computer software].

Kauffman, H. (2015). A review of predictive factors of student success in and satisfaction with online learning. *Research in Learning Technology, 23*, 26507. https://doi.org/10.3402/rlt.v23.26507

Kilgour, P., Reynaud, D., Northcote, M., McLoughlin, C., & Gosselin, K. P. (2019). Threshold concepts about online pedagogy for novice online teachers in higher education. *Higher*

Education Research & Development, 38(7), 1417-1431. https://doi.org/10.19173/irrodl.v14i1.1338

Kim, D., Wortham, S., Borowiec, K., Yatsu, D. K., Ha, S., Carroll, S., Wang, L., & Kim, J. (2021). Formative education online: Teaching the whole person during the global COVID-19 pandemic. *AERA Open, 7*(1), 1-12. https://doi.org/10.1177/23328584211015229

Kolvenbach, P. H. (2007). Cura personalis [Care of the whole person]. *Review of Ignatian Spirituality, 38*(1), 9–17.

Kuh, G. (2018). Whither holistic student development: It matters more today than ever. *Change: The Magazine of Higher Learning, 50*(3-4), 52-57. https://doi.org/10.1080/00091383.2018.1509590

Lowenthal, P. R., & Dunlap, J. C. (2018). Investigating students' perceptions of instructional strategies to establish social presence. *Distance Education, 39*(3), 281-298. https://doi.org/10.1080/01587919.2018.1476844

Martin, F., Budhrani, K., Kumar, S., & Ritzhaupt, A. (2019). Award-winning faculty online teaching practices: Roles and competencies. *Online Learning, 23*(1), 184–205. https://doi.org/10.24059/olj.v23i1.1329

Mathes, J. (2020, April 13). *A defining moment for online learning*. https://onlinelearningconsortium.org/a-defining-moment-for-online-learning/

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. J*ournal of Community Psychology, 14*(1), 6-23. https://doi.org/10.1002/1520 6629(198601)14:1%3C6::AID-JCOP2290140103%3E3.0.CO;2-I

Miller, K.E. (2021). A light in students' lives: K-12 teachers' experiences (re)building caring relationships during remote learning. *Online Learning, 25*(1), 115-134. https://doi.org/10.24059/olj.v25i1.2486

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

O'Malley, J. W. (2015). Jesuit schools of humanities: Yesterday and today. *Seminar on Jesuit Spirituality, 47*(1), 1–34. https://ejournals.bc.edu/index.php/jesuit/issue/view/641

Oswalt, S. B., Lederer, A. M., Chestnut-Steich, K., Day, C., Halbritter, A., & Ortiz, D. (2020). Trends in college students' mental health diagnoses and utilization of services, 2009–2015. *Journal of American College Health, 68*(1), 41-51. https://doi.org/10.1080/07448481.2018.1515748

Patton, L. D., Renn, K. A., Guido, F. M., & Quaye, S. J. (2016). *Student development in college: Theory, research, and practice*. John Wiley & Sons.

Quezada, R. L., Talbot, C., & Quezada-Parker, K. B. (2020). From bricks and mortar to remote teaching: a teacher education programme's response to COVID-19. *Journal of Education for Teaching*, 1-12.

Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business, 84*(2), 101–109. https://doi.org/10.3200/JOEB.84.2.101-109

Salmon, G. (2011). *E-moderating: The key to teaching and learning online* (3rd ed.). Routledge.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*. Psychometric Society.

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*(3), 543-553. https://doi.org/10.1016/j.compedu.2008.10.007

Vector Psychometric Group. (2020). *IRTPRO guide*. https://vpgcentral.com/wp-content/uploads/2020/06/IRTPROGuide.pdf

Vector Psychometric Group. (2022). *IRTPRO* (Version 6.0) [Computer software].

Wortham, S., Love-Jones, R., Peters, W., Morris, S., & Garcia-Huidobro, J. C. (2020). Educating for comprehensive wellbeing. *ECNU Review of Education, 3*(3), 406–436. https://doi.org/10.1177/2096531120928448

Zhai, Y., & Du, X. (2020). Addressing collegiate mental health amid COVID-19 pandemic. *Psychiatry Research, 288*, 113003.

Appendix
Formative Education Online Scale

When teaching online, how often do you do the following activities?

	Never [1]	Rarely [2]	Sometimes [3]	Often [4]	Very often [5]
I incorporate assignments that ask students to consider moral dilemmas.					
I encourage students to share their personal life experiences.					
I welcome students to contribute their spiritual or religious beliefs and values in classroom discussions.					
I provide opportunities in class for mindfulness and/or contemplation.					
I encourage students to incorporate what matters to them in the course assignments.					
I use material from my class to help students connect how their individual sense of purpose relates to serving other people.					
I design assignments and facilitate conversations that help students identify what is meaningful in their lives.					
I structure my courses to encourage students to think beyond their personal experiences and toward their ultimate contribution to a greater good.					
I encourage students to develop their own moral compass.					
My course attends to the integration of mind and spirit as a component of self-discovery.					

Figure A1

Calculating Raw Score

If the respondent answered at least 5 out of 10 items, then:

- For each "Never" response, assign a point value of "1."
- For each "Rarely" response, assign a point value of "2."
- For each "Sometimes" response, assign a point value of "3."
- For each "Often" response, assign a point value of "4."
- For each "Very often" response, assign a point value of "5."
- Sum all responses to obtain point total.
- Compute: $\frac{Point\ Total}{Items\ Answered} * 10$
- The raw score will range from 10 to 50.
- The raw score can be converted to an IRT Score that has been scaled to have a mean of 50 and a standard deviation of 10 (i.e., T-Score distribution) using Table A1.
- Round the raw score to the nearest whole number before converting to an IRT Score.

Table A1

Raw Score to IRT Score Conversion Table

Raw Score	IRT Score	Raw Score	IRT Score	Raw Score	IRT Score
10	18	25	41	40	56
11	22	26	42	41	57
12	24	27	43	42	58
13	26	28	44	43	59
14	27	29	45	44	61
15	29	30	45	45	62
16	30	31	46	46	64
17	32	32	47	47	65
18	33	33	48	48	67
19	34	34	49	49	69
20	35	35	50	50	73
21	36	36	51		
22	37	37	53		
23	38	38	54		
24	40	39	55		