Validation study of the *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* instrument using Rasch

Jennifer Mata-McMahon, Michael Haslip & Lance Kruse

View supplementary material

Published online: 12 Jun 2023.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Validation study of the *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* instrument using Rasch

Jennifer Mata-McMahon [a], Michael Haslip [b] and Lance Kruse [c]

aEarly Childhood Education, Sherman Center for Early Learning in Urban Communities, University of Maryland Baltimore County (UMBC), Baltimore, MD, USA; bEarly Childhood Education, Drexel University, Philadelphia, PA, USA; cChief Executive Officer, Invontics, Raleigh, NC, USA

ABSTRACT

Three validity evidence sources (test content, internal structure, and relationships to other variables) from the responses of 365 educators, purposely sampled from 36 US states, are presented to explore the use of the instrument, *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)*. Findings show expert panel agreed items accurately represented the desired construct and recommendations for revising multidimensional items were made by a psychometrician. The Rasch measurement analysis recommended collapsing the five-point frequency scale to four-point and removing two reverse-scored items. The revised instrument demonstrated excellent item fit, person and item reliability, separation, and practical unidimensionality. Relationships to other variables were established through no significant differences based on educator demographics. Differences were found based on school setting and educator values of spirituality, which aligned with expected differences. *ECE-SPC* is recommended as a self-report instrument, to determine to what extent early childhood educators nurture children's spirituality in secular educational settings.

Introduction

Children's spirituality has been studied globally by scholars of religious education for many decades (Berryman 1991; Larson and Keeley 2020; Ratcliff 2004; Roehlkepartian 2012; Sagberg 2015). A much newer topic of research is understanding children's spirituality from a secular perspective to learn how it might be supported and nurtured in public settings where religious education is not encouraged or even permitted. This line of inquiry states that the child can only be understood from a holistic perspective (J. P. Miller 2007, 2019; R. Miller 1988),

and that the spiritual domain is a part of each child (J. P. Miller et al. 2019) just as much as physical and cognitive domains. Therefore, the science of child development and early education, although non-religious in nature, can still consider how the spirit relates to holistic development and the impact that its nurturance or neglect may have on the child (Mata 2015).

To understand what is meant by 'secular spirituality' the term must be conceptualised independently from religion. Religion can be defined as a codified set of principles that a group of people believe in and adhere to in practice. These principles tend to be dogmatic, and the community of followers typically acts as guardians for adherence. In analysing what religion might mean in current times, Cox (2019), defines religion as 'identifiable communities that adhere to traditions that are transmitted from generation to generation with an overwhelming authority' (p. 331). Conversely, spirituality is an innate potential, a personal experience, in which connections are made within the self, with others, and with the Other, beyond the physical and material realm. Spirituality is experienced and expressed with an intention of understanding the meaning and purpose of life (Mata 2015). With this understanding of spirituality, the paths through which it can be experienced are expansive; spirituality can be cultivated through religious practice, yet it can also be developed in secular and non-religious ways (Elkins 1998). Focusing specifically on children, spirituality can also be experienced and expressed through '(a) joy, its expression, and its enjoyment; (b) concern for others through a display of kindness, compassion, and care; (c) the importance of relationships, and the value given to friends and family; and (d) the use and exploration of imagination' (Mata 2015, 103), encompassing artistic and musical activities in which children engage, as well as in ways they relate to nature.

Spirituality, understood as the capacity to connect and to better understand the purpose and meaning of life, has not been amply studied in the early childhood education field. Little is known about what early childhood educators might be doing in secular educational settings to support and nurture this very important aspect of child development (Mata 2012). With this in mind, Mata-McMahon et al. (2018, 2020) designed the first survey of its kind for national administration in the United States to explore if, how, and to what extent educators intentionally nurture spirituality in secular education settings. The instrument, accompanied by a manual, could be used by educators as a self-assessment tool, that would not only gauge what they are currently doing but also provide guidelines for future practices to make strides in nurturing spirituality for their students. Such an instrument would be an unprecedented tool for early childhood educators. The validation of the interpretation and use of this survey's scores, and its development into a measurement instrument for educators, is presented in this study.

Spirituality vs religion in early childhood education

In early childhood education, spirituality is often found in religious education literature, and within research studies typically conducted in private, religious settings. Teachers often believe that topics related to religion and spirituality need not be addressed in public school classrooms (Mata 2012), citing the separation between church and state in the first amendment to the constitution of the United States.

The first amendment to the United States Constitution states, 'Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof' (National Archives 2018), encompassing both what is known as the *establishment clause* and the *free exercise clause*. A common understanding of this portion of the first amendment of the U.S. constitution, referred to as the separation of church and state, is limited to the establishment clause, by which proselytising is not permitted in public settings funded by federal funds, such as public schools. Yet, if comprehensively understood, including the free exercise clause, the separation of church and state alerts that prohibition of the free exercise of religious beliefs should also not be enacted. Carpenter (2003) reminds us, 'we should remember that not only do the courts forbid any action by government schools not prompted by a "secular primary purpose" or which would "principally and primarily" aid religion; they also forbid any that would *inhibit* it' (p. 44). These clauses together permit multiple faith expressions in public schools or no faith expression. Seeing as the law is concerned with protecting the free expression or non-expression of religious identity, we interpret the constitution as permitting an understanding of spirituality that is not confined to any one religion, or religious group, and is a broader phenomenon.

Another impediment for teachers to nurture spirituality in the classroom is the typical way we define spirituality, often relating it to religion. Religion can be defined as a codified set of beliefs by which a community abides. It implies a doctrine, with rituals and practices that need to be adhered to and that are authoritatively transmitted from generation to generation (Cox 2019). Conversely, spirituality can be understood as an innate human potential, through which we connect to self, to others, and the Other beyond the materialist realm, a potential that can be nurtured to find meaning and purpose in life (Mata 2015). Spirituality can be developed and supported through religious practices, but it can also be nurtured through other paths. Elkins (1998), found eight distinct paths through which spirituality can be experienced, expressed, and developed. Through interviews with adolescents, he found that the feminine, the arts, the body, psychology, mythology, nature, relationships, and the 'dark nights of the soul' were all paths, different from religious practices, through which spiritual growth could be experienced.

Nurturing children's spirituality in public education settings is permissible because it can be explored independent of any particular religious community

or religious identity, as the field of psychology has shown, where spirituality is studied as a strength of character (Peterson and Seligman 2004). Therefore, the permissibility of spirituality in public schools depends not on questions about the separation of church and state but on questions of enriching child development: do practices associated with nurturing children's spirituality stimulate healthy, holistic growth and development? Early evidence suggests that it does (Mata-McMahon et al. 2018, 2020).

Factors influencing children's spirituality

Extant literature discusses many factors believed to nurture young children's spirituality. Much of this literature has been categorised into five major themes representing an emerging theoretical framework to understand young children's spirituality (Mata-McMahon et al. 2018). These themes include: 1) children's spirituality is nurtured with love and expressed as love, 2) children's spirituality includes forming close relationships and friendships, including a relationship with the earth, 3) children express and experience spirituality in *moments* of joy, wonder, and imagination, often associated with their engagement in expressive arts, 4) children's spirituality involves feeling part of something greater than oneself, often explored by asking big questions, and 5) children's spirituality is nurtured through the expression of virtues and values of character (e.g., kindness, fairness).

Nurturing spirituality in the curriculum has been conceptualised as assisting children to embrace the moment (Schein 2012), find meaning in the mundane (Bone, Cullen, and Loveridge 2007), create a connection between children and the earth, explore positive emotions in the present moment, engage in expressive activities through art and music (Mata 2015), and focus on the development of virtues in children (Mata-McMahon et al. 2018). Furthermore, a spiritually nurturing school environment deeply values the natural curiosity, sense of wonder, and joyful self-expression of children (Mata 2015; Mata-McMahon 2018), often through a play-based curriculum (Mata-McMahon 2019) that encourages children's humour (Schein 2017; Mata-McMahon 2017, 2019).

Nurturing children's spirituality in early education also focuses on providing children with an aesthetically pleasing environment (Montessori 1963; Steiner 1997; Bone and Fenton 2015). Children's need for a beautiful, stimulating environment has also been described by Reggio Emilia-inspired educators, although absent is any specific mention of spirituality (Gandini et al. 2015).

The characteristics associated with nurturing children's spirituality as mentioned above (e.g., creative expression, close relationships, virtue development) also promote holistic child development (Mata-McMahon et al. 2018, 2020). Therefore, the widely accepted framework for promoting children's holistic development in the early years, Developmentally Appropriate Practice (DAP), can be used as a framework for studying spirituality in early

childhood classrooms. DAP presents a guide to high-quality early education around the role of the teacher, the environment and curriculum, and the culture of the school (Bredekamp 2016; Copple and Bredekamp 2009). The instrument *Early Child Educators' Spiritual Practices in the Classroom (ECE-SPC)* has been organised similarly, with sections on educator's spirituality, curriculum, environment, and school culture.

The survey as administered had 54 questions organised in five sections: I) spiritual views, regarding beliefs held by the educator, II) activities and curriculum, for items dealing primarily with teaching strategies and techniques, III) classroom environment and schedule, IV) interactions and experiences, for actions taking place across the whole school and extending into the community, and V) demographic information. Thirty-six questions were presented as items with a five-point rating-scale, with five items in section I using a Likert-type agreement scale and 31 items in sections II, III, and IV using a frequency scale. Five questions were open-ended, asking educators to provide practical examples illustrating the rating-scale items. Thirteen questions were asked to gather respondents' demographic information.

To write items for sections II – IV the extant literature on children's spirituality (32 books and articles) was reviewed, particularly from a secular rather than a religious perspective, to identify educational activities that were aligned with nurturing the spirit (e.g., Gill and Thomson 2016; Bone and Fenton 2015; Minor and Grant 2014; J. P. Miller 2007). The procedure for item development followed this approach: 1) references to 32 publications were organised into a literature review table; 2) the literature was read and unpacked to identify specific suggestions about nurturing the spirit of children which were grouped into topic categories and subcategories (such as the category of connection with nature and the related subcategory of caring for plant and animal life); and 3) unique survey items were written to represent each category and subcategory of literature topics (e.g., times for playing in nature and children's caretaking of plant and animal life). These items were written as observable indicators of the concepts found in the body of literature. Table 1 shows a sample of how instrument items were developed for sections II – IV by mapping concepts in the literature to corresponding pedagogical implications and classroom environment characteristics. Scholars frequently discussed the same concepts (e.g., importance of nature; warm relationships) and therefore one survey item was often constructed to address the scholarship of multiple authors. Items were not borrowed from any other assessment tool.

Context of the study

The *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), &

Table 1. Sample of the instrument item development table.

Sample references	Sample concept	Sample instrument item	Instrument section[1]
J. P. Miller (2007) Woolley (2008)	Holistic development; develop interconnectedness	Use an integrated curriculum	(II) Activities & Curriculum
Steiner (1926/1982 Schein (2012)	Relationship between children and nature is promoted	Children play in nature	(III) Environment & Schedule
Montessori (1995) Fisher (1999)	Loving relationships ignite the spirit; quality relationships indicate spiritual health	Children show love for one another	(IV) Interactions & Experiences

[1]The survey has five sections. Section I of the survey asks about personal beliefs related to spirituality and Section V asks about respondents' demographic information; thus, those sections are not represented in this table.

National Council on Measurement in Education (NCME) 2014) provide one means to collect and evaluate the validity and reliability evidence of educational and psychological assessments. While there are various methods to construct validity arguments (e.g., Briggs 2004; Kane 2013; Mislevy 2004; Schilling 2004), many educational researchers are proponents of the *Standards* (AERA, APA & NCME 2014) to construct validity arguments to support the intended interpretation and use of test scores (see Goldstein and Flake 2016; Pellegrino, DiBello, and Goldman 2016; Sireci 2013). The *Standards* (AERA, APA & NCME 2014) identify five main sources of validity evidence: test content, response processes, internal structure, relationships to other variables, and consequences of testing.

Test content validity evidence refers to the degree that the items on the instrument accurately represent the desired construct (AERA, APA & NCME 2014). Such evidence is typically gathered through a careful review of items by a panel of subject matter experts to ensure the items align with the content domain. Response processes validity evidence suggests that individuals are completing the instrument in a manner consistent with the theorised construct (AERA et al., 2014). The use of cognitive interviews with a representative sample of participants is frequently used to obtain response processes validity evidence (Padilla and Benítez 2014). Internal structure is the degree to which the items and test components adhere to the desired construct, which is typically evaluated through a psychometric analysis (AERA et al., 2014). Relationships to other variables is another source of validity evidence, which explores how the test scores relate to a similar construct (i.e., convergent validity evidence) or a dissimilar construct (i.e., divergent validity evidence) (AERA et al., 2014). Lastly, consequences of testing validity evidence concerns the appropriateness of the proposed interpretation of test scores with their intended uses (AERA et al., 2014). For example, if an instrument is purported to provide insight into early childhood educators' nurturing of children's spirituality to encourage a change in teacher pedagogy, then it is important that educators use the results to make such changes in practice. While there are many methods to evaluate each source of validity evidence, the use of measurement theory provides insightful data to construct the validity argument.

Measurement theory can broadly be categorised as either Classical Test Theory (CTT) or modern measurement theory (Bond and Fox 2015). A primary concern with the use of CTT is that these methods find the sums of scores for the set of items and treat this total score as if it were interval-level data, which is a statistical assumption for most parametric analyses (Bond and Fox 2015). A second concern is that CTT suggests each item contributes equally to the measurement of the construct, no matter how challenging an item may be. As such, CTT is item dependent and does not allow for the comparison of instruments with varying numbers of items or with missing data (Wright 1996). Lastly, the reliability and standard error statistics are less generalisable in CTT. An instrument may demonstrate acceptable reliability statistics for one sample and poor reliability for another sample, which is particularly problematic for validation studies (Hambleton 2000). Such assumptions in CTT have resulted in decades of erroneous conclusions that have been widely reported across psychometrics (Bode and Wright 1999; Reise and Henson 2003; R. Smith 1996; Waugh and Chapman 2005). On the other hand, modern measurement involves the use of Item Response Theories (IRTs), which address many of the shortcomings associated with CTT.

Rasch measurement theory (Rasch 1960, 1980) is a type of item response theory (IRT) frequently used for evaluating an instrument's psychometric performance. See Table 2 for a brief comparison between Rasch, one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) IRT models.

The 1-, 2-, and 3-PL models include the parameters of item difficulty, discrimination, and pseudo-guessing. These parameters may be allowed to vary or be held constant, but these IRT models fit the model to the data and are thus descriptive in nature. Conversely, Rasch specifies there exists a relationship between an individual's ability and an item's difficulty, without other parameters, and is modelled as a probabilistic function (Wright and Stone 1979). The Rasch model is a unique type of IRT, which follows the Thurstonian conditions for measurement, including unidimensionality, linearity, invariance, and independence (Shaw 1991). Unidimensionality suggests that the instrument measures only one construct

Table 2. Comparison between Rasch, 1PL, 2PL, and 3PL-IRT models.

	Rasch	1PL	2PL	3PL
Data/Model Relationship	Fits data to the model (prescriptive)	Fits model to the data (descriptive)	Fits model to the data (descriptive)	Fits model to the data (descriptive)
Parameters	Item Difficulty	Item difficulty, Item discrimination (held constant), &Pseudo-guessing (held constant)	Item difficulty, Item discrimination, &Pseudo-guessing (held constant)	Item difficulty, Item discrimination, &Pseudo-guessing
Objective Measurement	Yes	No	No	No

(Wright and Masters 1982). Linearity infers a continuum from less to more along which the variable is measured (Wright and Masters 1982; Wright and Stone 1979). Invariance suggests that the unit of measure (i.e., logits, or log-odd-units) remains constant across different samples and along the continuum of the construct (Wright and Stone 1979). Lastly, independence refers to the item- and person-independence of the measure such that the measure is not altered by the object being measured (Wright and Masters 1982). Through meeting these four conditions, the Rasch model provides unique benefits for validation studies by separating the parameters (items from persons and persons from items) that is otherwise not possible in the other IRT models. As a result, the Rasch model has been frequently used within survey validation studies (e.g., O'Connor, Crawford, and Holder 2015; Sondergeld and Johnson 2014; Yang, He, and Liu 2018).

The present study utilises the *Standards* (AERA et al., 2014) as a basic framework for constructing the validity argument. Specifically, this study will focus on evaluating the evidence for test content, internal structure, and relationships to other variables. It is not uncommon for validation studies to focus on a selection of some of the sources of validity evidence (Sondergeld, 2020). A single study should not be considered the sole validation study as robust validity arguments are the culmination of various claims and different sources of evidence across multiple studies (Krupa, Bostic, and Shih 2020). As such, this study will primarily focus on the psychometric data to evaluate three of the five sources of validity evidence, while future research is suggested to explore the remaining sources through primarily qualitative inquiries. The psychometric data were produced using the Rasch measurement model (Rasch 1960, 1980) given its utility for measuring unidimensional constructs with invariant units of measure (i.e., logits, or log-odd-units) that are sample and item independent. Such characteristics result in more conservative and generalisable statistics than those generated from CTT, which aligns with the intended purpose of a validation study.

Research questions

The main purpose of this study is to determine if the statistical evidence for the *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* instrument accurately measures how and to what extent early childhood educators nurture children's spirituality in secular educational settings. The guiding research question posed to evaluate the validity evidence for this instrument stems directly from three of the five main sources of validity evidence proposed by the *Standards* (AERA et al., 2014), and the selection of the Rasch measurement model (Rasch 1960, 1980) for psychometric analysis. Thus, the guiding research question posed for this study is:

(1) Does the *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* instrument demonstrate acceptable validity evidence regarding test content, internal structure, and relationship with other variables, in its

measurement of how and to what extent early childhood educators nurture children's spirituality in secular educational settings?

Methods

Respondents

The population targeted for this study were U.S. early childhood educators working in secular educational settings with children between ages zero and eight. A comprehensive list of approximately 10,500 emails sourced from all 50 states' educational department databases was compiled for survey distribution, collecting 365 responses which comprised the sample.

Demographic characteristics of the study respondents are presented in Table 3. Most respondents were female (94%), white (77%), held a college degree (76%), worked in a teaching role (72%) versus administrative role (28%), had at least 8 years of teaching experience (78%), and were between 33–60 years of age (74%). Respondents were relatively balanced between those who work in public secular (46%) and private secular (43%) settings, and between locations in urban (35%), suburban (38%) and rural (27%) regions of the country. Respondents came from 36 U.S. states, the District of Columbia, and the U.S. territory of Puerto Rico. Most respondents came from Pennsylvania (59%) where the researchers had access to a state-wide database of early childhood centres, followed by Maryland (6%), Alaska (6%), Texas (2%), and the District of Columbia (2%). Other states represented less than 2% of the sample. The demographic characteristics of our sample are similar to national averages regarding the early childhood education workforce in terms of gender differences, race/ethnicity, work setting (urban, suburban, rural), and the economic background of families served (National Workforce Registry Alliance n.d.; Parker et al. 2018; Semega et al. 2020; U.S. Bureau of Labor Statistics 2021).

We also asked survey respondents to identify generalised demographic characteristics of the children they serve (Table 4). Overall, children were relatively balanced across low, middle, and high-income categories. Settings served mostly minoritized children or an even mix between white and minoritized children, in 54% of the settings represented, while 46% of the settings had mostly white children.

Instrument design

The instrument *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* was originally designed by Mata-McMahon, Haslip and Schein in 2016 as a survey, using the Process Model for Assessment Design/Selection and Validation by Chatterji (2003). The initial purpose of the instrument design was to survey how and to what extent early childhood educators, working with children between ages zero and eight, in secular educational environments, were nurturing children's spirituality. The final outcome will be an

Table 3. Demographic characteristics of respondents ($N =$ 363).

Characteristic	n	%
Age ($n = 355$)		
22–32	56	16
33–46	143	40
47–60	121	34
61 and above	35	10
Gender ($n = 361$)		
Female	338	94
Male	23	6
Race/Ethnicity ($n = 355$)		
Asian (Asian, Middle Eastern, Indian)	8	2
Bi-racial	3	<1
Black or African American	52	15
Hispanic or Latinx	18	5
Native American, Native Hawaiian or Pacific Islander	2	<1
White	272	77
Highest Education Completed ($n = 362$)		
High school diploma/GED	13	4
Some community college or CDA	36	10
Associates degree	36	10
Bachelor's degree	60	17
Bachelor's degree + teaching licence	55	15
Master's degree	148	40
Doctoral degree	14	4
Job Title ($n = 348$)		
Lead teacher	150	43
Assistant or co-teacher	13	4
Specialist (reading, maths, music, P.E., science) or special education teacher	15	4
Homecare provider or caregiver	66	19
Administrator (centre director, owner, family childcare provider)	99	28
Other	5	1
Years of Teaching Experience ($n = 351$)		
0–3	34	10
4–10	91	26
11–20	125	36
21 or more	101	28
Work Setting ($n = 363$ valid)		
Public secular	165	46
Private secular	157	43
Religiously affiliated	41	11
Region of Work Setting ($n = 360$)		
Urban	127	35
Suburban	136	38
Rural	97	27

instrument with scores that have been validated as a self-assessment measure of the extent to which early childhood educators nurture children's spirituality in the classroom, to be provided with guidance on how to improve these practices, if so desired.

Chatterji's (2003) Process Model involved, in phase I, the specification of the target population, purpose, and construct to be measured. In phases II, III and IV, it involved the completion of an iterative process, including developing the instrument's assessment specifications, developing the instrument's assessment tool, and conducting a content validation utilising the assessment tool

Table 4. Child demographics.

Characteristic	n	%
Child Demographics (*n* = 359)		
Most white children	165	46
Most minority children	75	21
White and minority children (even mix)	119	33
Child SES (*n* = 360)		
Most high income	16	4
High and middle income	67	18.5
Most middle income	85	23.5
Low and middle income	110	31
Most low income	82	23

(explained below in the Data Analysis section, under Expert Panel Review). Lastly, it included an empirical validation, which is the study presented in this paper. The design of the *ECE-SPC* stemmed directly from the literature on the construct to be assessed: how children's spirituality can be nurtured in secular educational settings.

The literature on children's spirituality in early childhood education is rich with religious education curricula and program implementation, in the US and abroad, mainly in Canada, the UK, and Australia (Allen 2008; Espinoza, Estep, and Morgenthaler 2018; Erricker, Ota, and Erricker 2001; Larson and Keeley 2020; Lawson, 2012; Lawson and May 2019; Ota and Erricker 2005; Tolbert 2014). Conversely, there are fewer publications regarding nurturing children's spirituality in secular educational settings. Nevertheless, a review of the literature was conducted to establish which recommendations were suggested for early childhood educators looking to nurture children's spirituality and how to further support educators in facilitating these secular spiritual practices in the classroom, to determine which specific aspects of children's spirituality should be included in the instrument.

Knowing that early childhood educators' views and understandings of spirituality may guide how they design and deliver curricula, it was decided to begin the survey by asking about educators' spiritual views. Educators' spiritual views were measured in Section I: Spiritual Views and Practices, comprising five Likert-type items and two open-ended questions. These items were not intended to measure the overarching construct but served two purposes. First, the items served as a quality check on survey responses as item three and four require answers on opposite ends of the scale to be aligned. Second, answers to these items were used to categorise respondents into groups to allow for comparisons in survey scores as the relationships to other variables source of validity evidence. To document the specific teaching strategies and pedagogical techniques the educators were using to nurture spirituality, Section II: Activities and Curriculum was included, comprising 17 items using a five-point frequency scale and one open-ended question. The organisation of the classroom environment and the schedule developed by teachers comprised Section III: Classroom Environment and Schedule, with six items using a five-point frequency scale and one open-ended question. Section IV:

Table 5. Composition of the original ECE-SPC instrument.

Instrument Section	Construct	Rating scale items	Scale	Questions
Section I: Spiritual Views and Practices	Definitions and practices of spirituality for educators	5 (items 1, 2, 3, 4, 5)	Endorsement	2 open-ended (questions 6, 7)
Section II: Activities and Curriculum (teaching strategies and techniques)	Pedagogical design for nurturing children's spirituality	17 (items 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24)	Frequency	1 open-ended (question 25)
Section III: Classroom Environment and Schedule	Environment and schedule intentionally set to nurture children's spirituality	6 (items 26, 27, 28, 29, 30, 31)	Frequency	1 open-ended (question 32)
Section IV: Interactions and Experiences (in school and around the community)	School-wide interactions and community experiences to foster children's spirituality	8 (items 33, 34, 35, 36, 37, 38, 39, 40)	Frequency	1 open-ended (question 41)
Section V: Demographic Information	Respondents' demographic characteristics	-	-	13 multiple choice (questions 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54)

Interactions and Experiences, with eight items using a five-point frequency scale and one open-ended question, was developed to measure how the overall school and surrounding community was being used by educators to support children's spirituality. Lastly, Section V: Demographic Information, was included with 13 multiple choice questions, to determine the demographic characteristics of the early childhood educators surveyed, understand the children they work with, and the specifics of their educational settings.

Overall, the *ECE-SPC* instrument has five sections, comprising 54 questions, of which 36 are rating-scale items (i.e., a Likert-type endorsement scale, and a five-point frequency scale), five are open-ended questions, and 13 are multiple choice questions related to demographics. The original instrument designed as a survey can be seen in Supplementary A. A summary of the composition of the original *ECE-SPC* instrument can be seen in Table 5, below.

Validity framework

The present study utilises the sources of validity evidence outlined in the *Standards* (AERA et al., 2014) to construct a validity argument for the *ECE-SPC*. Specifically, the *ECE-SPC* was designed to measure early childhood educators' nurturing of children's spirituality in secular school settings. As such, a variety of validity evidence is collected to evaluate the interpretation and use of the instrument scores to support this intended use (AERA et al., 2014). The specific sources of validity evidence and associated data sources are identified in Table 6.

Table 6. Source of validity evidence and its associated data sources.

Validity Evidence	Data Source
Test Content	Expert Panel Review
	Rasch Wright Map
Internal Structure	Rasch Psychometric Analysis:
	item/person reliability
	item/person separation
	item/person fit
	item unidimensionality
	item/person targeting
Relationships to Other Variables[1]	Demographic Variables
	Educator Race/Ethnicity
	Type of Educator
	Education Level
	Years of Experience
	School Variables
	School Setting[2]
	School Region
	Demographic of Students
	SES of Students
	Values of Spirituality
	Important Personally[3]
	Important for Children[3]
	Important in Schools[3]
	Only at Home[4]
	Domain of Life

Note: [1]It was expected that no significant differences would be found based on the listed variables, except where indicated.
[2]Expected difference such that educators from private schools have higher scores than those from public schools.
[3]Expected difference such that educators who agree with these values of spirituality have higher scores than those that disagree.
[4]Expected difference such that educators who disagree with this value of spirituality have higher scores than those that agree.

Data analysis

Expert panel review

The expert panel review was conducted through a content validation process completed by reaching out to three experts in the field of children's spirituality. A content validation instrument assessment tool was developed to record the expert's evaluation of the *ECE-SPC*. A cover letter was drafted; an email was sent to each expert including the *ECE-SPC* (survey version) and the instrument assessment tool for use when evaluating the *ECE-SPC*. Of the three experts contacted, two responded and completed the evaluation of the *ECE-SPC* by filling in and returning the content validation instrument assessment tool. Even though six experts is the most common number recommended for content validation, two experts is concurred as the minimum acceptable (Yusoff 2019), thus the expert responses received were considered satisfactory.

A psychometrician was also consulted to evaluate the overall instrument design, as well as item redundancy and/or multidimensionality.

Rasch psychometric analysis

The Rasch psychometric analysis followed an iterative process of reviewing the frequency rating scale performance, person fit, and item fit to ensure the data appropriately fit the Rasch model. The analysis was completed using Winsteps Version 4.0.1 (Linacre 2016). The first analysis involved rating scale performance.

Rating scale performance. The five-point frequency rating scale was first evaluated to ensure participants were meaningfully interpreting each scale category in a similar and consistent manner. The specific guidelines for evaluating rating scale performance according to Linacre (2002) are as follows:

(1) There should be at least 10 observations in each rating scale category to improve the estimate and stability of step calibrations.
(2) Average measures should advance monotonically such that higher rating scale categories (i.e., several times each day) are produced by higher measures (i.e., more nurturing of children's spirituality) and lower rating scale categories (i.e., once a month or less) are produced by lower measures (i.e., less nurturing of children's spirituality).
(3) Outfit mean-squares should be less than 2.0, which suggests there is minimal unexplained noise (or randomness) in the model.
(4) Step calibrations for a five-category scale must advance by at least 1.0 logits but less than 5.0 logits, which indicates the rating scale meaningfully differentiates between two different ideas (e.g., between 'About once a week' and 'Several times a week').
(5) The probability curve for each scale should have a distinct peak to indicate the category is meaningfully measuring the construct.

After the rating scale was determined to be functioning appropriately, then the item and person reliability and separation were evaluated.

Item and person reliability and separation. The Rasch analysis estimates reliability for both items and persons, which follows the interpretation of traditional measures of internal consistency in CTT (e.g., Cronbach alpha) such that a reliability closer to 1 indicates more internal consistency in the measure (Wright and Masters 1982). However, the Rasch reliability is more conservative than Cronbach alpha as traditional measures of reliability are calculated from the raw data, which is negatively impacted by its nonlinearity (Wright and Masters 1982). Separation is also provided, which indicates the spread of items and persons, such that the larger the spread, the easier it is to meaningfully differentiate items and persons from each other (Wright and Stone 2004). The following guidelines assist in the evaluation of item and person reliability and separation: acceptable = 1.50 separation and 0.70 reliability; good = 2.00 separation and 0.80 reliability; and excellent = 3.00 separation and 0.90 reliability (Wright and Masters 1982; Sondergeld and Johnson 2014). When

making any changes to the data, the item and person reliability and separation indices should be re-estimated and compared to previous iterations to ensure that any changes have improved (or at least not hindered) the performance of the instrument (Boone, Staver, and Yale 2014). The primary method of improving item and person reliability and separation is through evaluating item and person fit.

Item and person fit. When implementing the Rasch model, the data must appropriately fit the model. Rasch fit statistics assist to understand how the observed responses to items by persons compare to the model's expected response (E. Smith et al. 2002). Item fit statistics provide insight into which items may not be assessing the same construct, while person fit statistics provide insight into which persons may be responding irregularly to items (E. Smith et al. 2002). There are five statistics used to evaluate the fit of items and persons: point biserial, outfit/infit mean-square, and outfit/infit z-standardised.

Point biserial is the correlation between the response to an item and the overall measure score. All point biserial should be positive, otherwise a negative point biserial suggests the item or person contradicts the measurement of the desired construct (Boone, Staver, and Yale 2014; Wright and Stone 2004). Two fit indices are provided: outfit and infit. Outfit statistics are sensitive to unexpected responses whereas infit statistics are sensitive to responses near a given item difficulty or person ability (Boone, Staver, and Yale 2014; Wright and Stone 2004). It is recommended to focus on outfit statistics as these indicate the prevalence of misinformation that may hinder the measurement of the construct (Linacre and Wright 2012). Fit statistics are primarily reported through a mean-square, which is a chi-square calculation with a mean of 1.0 logits (Boone, Staver, and Yale 2014). Lower values indicate overfit (i.e., the model overpredicts the data) and higher values indicate underfit (i.e., there is too much misinformation in the data) (Boone, Staver, and Yale 2014). Items and persons with appropriate fit are within a range of 0.6 to 1.4 logits (Wright & Linacre, 1994). If an item or person is outside of the suggested range, then it's associated z-standardised value should be investigated. If the z-standardised value is larger than 2.0, then the item or person with poor fit should be investigated. Items and persons with negative point biserial values and/or infit/outfit statistics outside of the acceptable range should be removed from the analysis to create a more parsimonious measure of the construct (Boone, Staver, and Yale 2014).

Unidimensionality. The investigation of unidimensionality first involved the evaluation of item fit. Specifically, items with negative point-biserial values or poor fit were removed from the analysis, with the remaining items working cohesively together to measure the desired construct. However, a more nuanced investigation of dimensionality involved the Rasch Factor Analysis, which identifies the common variance that is not modelled by the Rasch measure and indicates the potential presence of a secondary construct in the measure (Wright and Stone 2004). Recommendations for identifying

a secondary construct are not explicit but explained variance should be greater than 50% for a unidimensional construct (Linacre 2003). Conversely, if the variance in the first contrast is greater than 10% of the total variance, there may be evidence of a secondary construct (Linacre 2003). In such a case, a qualitative review of items is necessary in order to determine if the cluster of items are theoretically a different construct or if they may be considered as part of a practically unidimensional construct (Linacre 2003).

Wright map. Lastly, the Wright Map represents the continuum from more to less of the construct with persons and items plotted on the map based on their person measure score and item difficulty. The Wright Map allows for a visualisation of the construct, which may assist in evaluating the test content validity evidence (i.e., are the expected difficult items at the top of the map and the expected easier items at the bottom?). Additionally, the Wright Map provides insight into the overall performance of the instrument by determining a relatively normal distribution of persons, equally spaced items from high to low, approximately equal person measure mean and item difficulty mean, and items that are beyond the highest and lowest person (Boone, Staver, and Yale 2014; Wright and Masters 1982).

Statistical analyses

To evaluate differences based on educators' demographic variables, school characteristics, and values of spirituality, a series of one-way Analysis of Variances (ANOVAs) were conducted for all categorical variables with the Rasch person measure score being the interval-level dependent variable (see Table 6 for expected differences based on these analyses). The five-point Likert-type scale options of response for each of the values of spirituality questions were recorded to be either 'Agree' (by combining 'Strongly Agree' and "Agree), 'Neutral', or 'Disagree' (by combining 'Strongly Disagree' and 'Disagree'). Also, due to low numbers of racial minority educators, only educators who reported being white or Black/African American could be compared. It was decided not to collapse all racial minority educators together because there is no evidence to suggest that non-white educators are more similar to each other regarding nurturing of spirituality than they are with white educators. Post-hoc analyses were conducted for tests with significant omnibus results ($p < 0.05$) using Fisher's Least Squared Distance (LSD), which allows for pairwise comparisons with unequal groups (Fisher 1936). Effect sizes for the ANOVAs were computed using η^2 such that 0.01 is small, 0.06 is medium, and 0.14 is large (Cohen, 1988). To evaluate the correlation between educators' years of experience (interval-level) and their Rasch person measure score (interval-level) a Pearson correlation was conducted. Effect size for the Pearson correlation was computed using r^2 such that 0.1 is small, 0.3 is medium, and 0.5 is large (Cohen, 1988). Statistical

tests were conducted using SPSS version 27 and were interpreted at an alpha level of 0.05.

Findings

The findings will be presented in the following sections organised by the various data sources for each type of validity evidence. Note, the instrument items are discussed by referencing their section number followed by their unique item number (e.g., 2.9, is item nine in section two).

Test content validity evidence

The content validity evidence follows the Expert Panel Review (see Data Analysis). Expert number one rated all items as assessing the construct for which they were developed to a great extent, with one exception. A suggestion was made to modify one item (current item 4.39) to make it more explicit and less general. This modification was made. Regarding the open-ended questions, no suggestions were made for modifications.

Expert two rated most items as assessing the construct for which they were developed to a great extent or a moderate extent. Regarding the open-ended questions, recommendations on the wording for three questions (current questions 1.6, 1.7, and 2.25) were made by expert two, and revised in the final instrument.

The psychometrician identified six multidimensional items in the instrument and offered suggestions to modify them. The items were revised, and the instrument was expanded from a total of 25 items to a new total of 36 items, as included in the original survey version of the *ECE-SPC* (see Supplementary A). The psychometrician also made suggestions to incorporate questions regarding the demographic information of the children being served by the educators being studied and those questions were added to the instrument. She also suggested moving the demographic section from the beginning to the end of the survey and this was done.

Internal structure validity evidence

The initial Rasch analysis of the *ECE-SPC* data included all 362 respondents and the 31 frequency rating scale items that measured educator's nurturing of children's spirituality (the five values of spirituality items, Section I, were not included in the psychometric analysis as they were not designed to measure the same construct on the same scale). Those items and respondents demonstrated excellent person and item reliability and separation (see Table 7). The results of the iterative Rasch analysis processes are described in the following sections.

Table 7. ECE-SPC summary statistics and performance.

Rasch Index	Instrument Version		
	Five-Point Frequency Scale 362 Respondents 31 Items	Four-Point Frequency Scale 362 Respondents 29 Items	Four-Point Frequency Scale 338 Respondents 29 Items
Person Reliability	0.90 (Excellent)	0.92 (Excellent)	0.92 (Excellent)
Person Separation	3.08 (Excellent)	3.34 (Excellent)	3.36 (Excellent)
Item Reliability	0.99 (Excellent)	0.99 (Excellent)	0.99 (Excellent)
Item Separation	11.50 (Excellent)	11.97 (Excellent)	12.53 (Excellent)

Rating scale performance

Each frequency rating scale category contained more than 10 counts, the category measures advanced monotonically, and the outfit mean-square values were less than 2.0. However, the step calibrations between 'About once a week' and 'Several times a week' (0.66) and between 'Several times a week' and 'Daily' (0.86) were less than the 1.0 threshold (Linacre 2002). Additionally, the probability curves for 'About once a week' (i.e., #2) and 'Several times a week' (i.e., #3) did not demonstrate distinct peaks (see Figure 1), which suggests that participants were not meaningfully differentiating between these rating scale options.

Given the small step calibrations and non-distinct probability curve peaks, the second and third options were merged, creating a four-point frequency rating scale (i.e., once a month or less, a few times a week, daily, and several times each day). With the four-point rating scale, the instrument satisfied all rating scale performance criteria, including step calibrations greater than 1.0 and distinct probability curve peaks (see Figure 2).

Item and person fit

After revising the frequency rating scale, all of the items, except for two, demonstrated acceptable fit (0.6 to 1.40 logits). The two misfitting items (4.39 and 2.17) had high outfit mean-square values (1.85 and 1.59, respectively) along with high z-standardised values (9.6 and 7.0, respectively). In reviewing these items, these were the only two items on the instrument that required reverse coding. That is, these two items were phrased such that an educator who has more nurturing of children's spirituality would report a lower frequency of doing those activities. The poor fit statistics indicate that respondents were not consistently responding to these reverse-scored items and thus these two items were removed from future analysis. The remaining items on the instrument (4-point frequency scale, 362 respondents, and 29 items) demonstrated acceptable fit (between 0.60 and 1.40 logits) and slightly improved the item and person reliability and separation indices (see Table 7).

Regarding person fit, there were 13 respondents who demonstrated negative point biserial values and 11 respondents had very high outfit
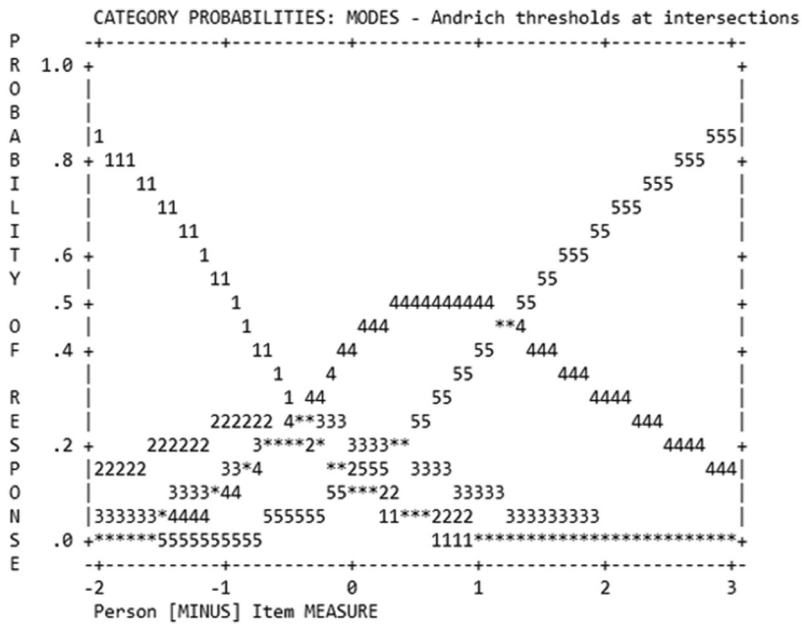
Figure 1. Probability curves for five-point frequency rating scale with 365 respondents and 31 items. Rating scale option 2 (about once a week) and 3 (several times a week) do not demonstrate distinct peaks, indicating a lack of differentiation by respondents.

mean-squares (>2.00) and z-standardised values (>6.0). Since these 24 respondents were contributing more misinformation than information to the model, they were removed from the analysis. Further analysis found that these 24 respondents were representative of the sample and did not significantly alter the demographics of the respondents. The remaining 338 respondents demonstrated acceptable fit with no significant mean-square values outside of the recommended thresholds. The revised instrument (4-point frequency scale with 338 respondents and 29 items) demonstrated slightly improved item and person reliability and separation indices (see Table 7).

Unidimensionality

The Rasch PCA results from the revised instrument indicated that the model accounted for 52.5% of the total variance and the first contrast contained only 5.5% of the total variance (Linacre 2003). Since the model accounted for more than 50% of the total variance and the first contrast less than 10% of the total variance, along with the acceptable item fit indices, the instrument was determined to be practically unidimensional.

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
 P   -+---------+---------+---------+---------+---------+---------+-
 R  1.0 +                                                             +
 O     |                                                             |
 B     |                                                             |
 A     |111                                                        4|
 B   .8 +    111                                              444 +
 I     |       11                                            44     |
 L     |        11                                         44       |
 I     |         11                                      44         |
 T   .6 +          11                                   44          +
 Y     |           1                                  44           |
     .5 +            11              3333333333  44               +
 O     |             11  222      333          **3                |
 F   .4 +           222*2   2222*3        4     333               +
       |           222    11    33 222    44       33             |
 R     |          222      1 33    22   44       333              |
 E     |         222        3*1       2244        333             |
 S   .2 +    222            33   11     44222                 333 +
 P     |2222             333       111 444     222              3|
 O     |            3333          44*11         2222            |
 N     |         333333      444444    111111     2222222       |
 S   .0 +*******444444444444444          11111111111111********+
 E     -+---------+---------+---------+---------+---------+---------+-
        -3        -2        -1         0         1         2         3
        Person [MINUS] Item MEASURE
```
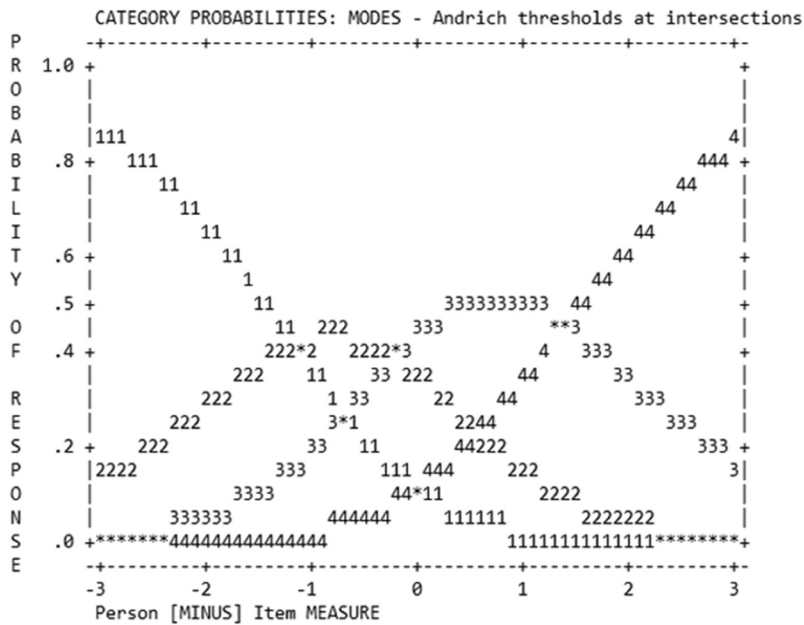
Figure 2. Probability curves for four-point frequency rating scale with 362 respondents and 31 items. The merged second and third rating scale options improved the rating scale performance.

Wright map

From the Wright Map (see Figure 3), it appears there was a relatively normal distribution of respondents with equally spaced items. However, there were not enough difficult or easy items to fully span the ability range of the sample. Additionally, the mean item measure (always 0.00 logits) was slightly lower than the mean person measure (0.57 logits), however it was within two standard errors of the mean item measure, which suggests the items are at a similar difficulty level as the ability of the sample. Although the instrument could benefit from both easier and more difficult items, the instrument functions well psychometrically.

The Wright Map also provided further understanding of the construct. The easiest items to endorse addressed paying attention to children when they have an interest/question (2.14), commenting on spontaneous discoveries of children (2.15), showing love to others (3.36), and demonstrating empathy for others (3.35). On the other hand, the hardest items to endorse addressed daily yoga (2.20), daily meditation (2.21), and sharing experiences with the spiritual realm with others (3.34). These findings are consistent with the literature, which shows that teachers seem to engage in conversations with children when they show interest in spirituality or ask questions related to spirituality (e.g., where do we
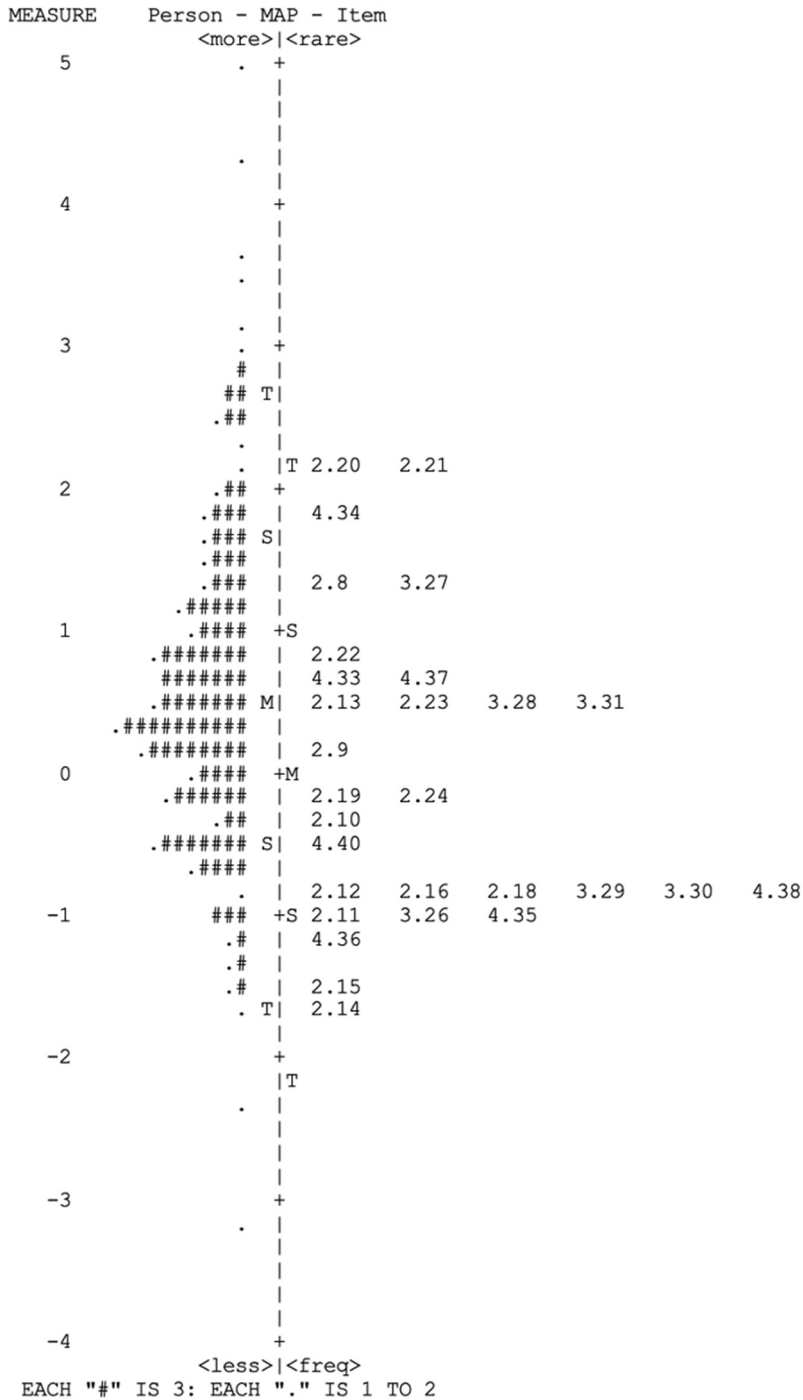
```
MEASURE     Person - MAP - Item
                <more>|<rare>
   5              .  +
                     |
                     |
                     |
                  .  |
                     |
   4                 +
                     |
                  .  |
                  .  |
                     |
                  .  |
   3              .  +
                  #  |
                 ## T|
                .##  |
                  .  |
                  .  |T 2.20    2.21
   2            .##  +
               .###  |   4.34
               .### S|
               .###  |
               .###  |   2.8     3.27
              .####  |
   1           .####  +S
            .######  |   2.22
             ######  |   4.33    4.37
            .###### M|   2.13    2.23    3.28    3.31
          .#########  |
            .########  |   2.9
   0           .####  +M
             .#####  |   2.19    2.24
               .##  |   2.10
            .####### S|   4.40
              .####  |
                  .  |   2.12    2.16    2.18    3.29    3.30    4.38
  -1              ###  +S 2.11    3.26    4.35
                 .#  |   4.36
                 .#  |
                 .#  |   2.15
                  . T|   2.14
                     |
  -2                 +
                     |T
                  .  |
                     |
                     |
                     |
  -3                 +
                  .  |
                     |
                     |
                     |
                     |
  -4                 +
                <less>|<freq>
     EACH "#" IS 3: EACH "." IS 1 TO 2
```

Figure 3. Wright map for four-point frequency scale with 338 respondents and 29 items. Each "#" represents three respondents and each "." represents one-two respondents. Easier items and less able persons are at the bottom while more difficult and higher able persons are at the top.

go when we die?) (Tomlinson et al. 2016) more frequently than they promote the practice of contemplative practices such as yoga or meditation (Hart 2003).

Relationships to other variables

The One-way ANOVA results for each categorical independent variable are presented in Table 8. The findings align with the hypothesised results such that no significant differences were observed for any of the variables except for school setting and four of the five values of spirituality variables. Effect sizes ranged from small to medium (0.034 to 0.110) with the independent variables accounting for 3.4% to 11% of the variance in the person measure score. Regarding the continuous independent variable, the correlation between educators' years of experience and their person measure scores was weak ($r = 0.193$, $p < 0.001$) with only 3.73% of the variance in person measure scores being accounted for by their years of experience ($r^2 = 0.0373$). This finding is suggested by the literature, as nurturing spirituality is not necessarily understood as a hierarchical construct that can be expected to increase in an upward linear direction as educators augment their technical expertise with time (J. P. Miller 2000). This can be seen in findings from a case study conducted in the Ivory Coast (Gottlieb 1998) in which infants were reported by the indigenous Beng community as being more spiritual than adults since they were closer to the spiritual realm. Beng adults reported that infants lead profoundly spiritual lives and, 'in fact, the younger they are, the more thoroughly spiritual their existence is said to be ... [thus]

Table 8. One-way ANOVA results for each categorical independent variable with person measure score as the dependent variable.

Category Variable[1]	F-statistic	η^2	Significant Pairwise Comparisons
Educator Demographics			
Educator Race/ Ethnicity	2.191	0.007	N/A
Type of Educator	1.391	0.023	N/A
Education Level	1.645	0.031	N/A
School Variables			
School Setting	22.354***	0.110	Private vs Public (p<.001)Private vs Religious (p=.010)Public vs Religious (p<.001)
School Region	0.465	0.003	N/A
Demographics of Students	2.107	0.012	N/A
SES of Students	2.313	0.025	N/A
Spirituality Views and Practices			
Important Personally	8.605***	0.046	Agree vs Disagree (p=.030)
Important for Children	6.258**	0.034	Agree vs Disagree (p=.025)Neutral vs Disagree (p=.003)
Important in School	8.879***	0.047	Agree vs Disagree (p<.001)Neutral vs Disagree (p=.010)
Only at Home	5.784**	0.031	Agree vs Disagree (p=.001)
Domain of Life	1.804	0.10	N/A

Note: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
[1]See demographics table for the options associated with each variable.

it follows that newborns are considered the most spiritual of all living humans' (Gottlieb 2006, 151). This is contrary to a western belief, in which adults might be considered more spiritual than children since they have been developing spiritually for a longer time than children. This non-western understanding of spirituality as non-linear in its development could help explain why less experienced educators could also be found to be more willing to support children's spirituality than those with more years of experience, as experience can be understood as unrelated to the extent to which educators nurture children's spirituality.

ANOVA post hoc results identified significant pairwise comparisons between each school setting such that educators from religious schools had significantly higher scores for nurturing spirituality ($M = 1.21$, SD = 1.27) than those from private secular schools ($M = 0.74$, SD = 1.09) and those from public schools ($M = 0.160$, SD = 0.917), and with those from private secular schools significantly higher than those from public schools. This is not surprising given the common misunderstanding given to the separation of church and state (Carpenter 2003) and the common association of the definition of spirituality with religion (Cox 2019; Mata 2015). Consequently, teachers tend to shy away from including topics related to spirituality in the public classroom.

Similarly, it was found that educators who agreed with spiritual beliefs questions as *Important Personally* ($M = 0.63$, SD = 1.09), *Important for Children* ($M = 0.64$, SD = 1.10), and *Important in School* ($M = 0.71$, SD = 1.09) had significantly higher average scores than those who disagreed with those items ($M = 0.19$, SD = 0.98; $M = 0.09$, SD = 0.69; and $M = 0.08$, SD = 1.07; respectively). Educators who agreed with *Only at Home* had significantly lower scores ($M = 0.22$, SD = 1.07) than those who disagreed with it ($M = 0.68$, SD = 1.12). Such findings align with the hypothesised differences between groups.

Discussion

The purpose of this study was to evaluate the validity evidence to support the interpretation and use of the *ECE-SPC* to measure how and to what extent early childhood educators nurture children's spirituality in secular educational settings. Through the exploration of test content, internal structure, and relationships to other variables (AERA et al., 2014), there is sufficient validity evidence to support the use of the *ECE-SPC* for its intended purpose. Specifically, the expert panel agrees the items reflect the domains identified from the literature and the Rasch analysis suggests the instrument performs psychometrically well, by providing meaningful measures of educators' nurturing of children's spirituality. Relationships to other variables also confirmed the expected relationships between respondent's values of spirituality and their level of nurturing spirituality.

From the analysis of spirituality belief items, educators who agreed that spirituality is important to them personally, important for children, and important at school were significantly more likely to have higher scores on the *ECE-SPC* than those who disagreed. Conversely, educators who believed spirituality should only be fostered at home were significantly more likely to score lower than those who disagreed. These findings align with existing research suggesting that educators' personal beliefs impact their practice in the classroom (Mansour 2009; Pajares 1992; Thomson and Nietfeld 2016). Additionally, the survey results were similar between classroom teachers and directors/administrators in the sample (e.g., childcare centre directors), yet further research exploring the use of the instrument amongst broader director/administrator populations is necessary (i.e., elementary school principals). While this type of research is one of the first conducted, these findings are intriguing as the field continues to better understand how early childhood educators nurture spirituality in children in secular settings. Additionally, the Rasch analysis allows for nuanced insight into what types of behaviours are more or less difficult for educators to endorse (see Table 9).

Table 9 orders the endorsement difficulty level of the instrument items from easiest (pay attention to children's questions and interests) to hardest (meditation is included in daily activities). Six difficulty level clusters were identified and have been labelled by the researchers. These clusters can be interpreted as domains of practice within spiritually nurturing education. Results imply that it would be easiest for teachers to begin a spiritually informed pedagogy by incorporating practices related to 'spiritual responsiveness' followed by 'promoting spiritual growth through the environment'.

Existing literature provides a rationale for these endorsement difficulty levels. For example, being 'responsive' is a foundational component of quality teacher-child interactions (La Paro, Pianta, and Stuhlman 2004) and has been frequently taught to pre-service and in-service early childhood educators using texts such as *101 Principles for Positive Guidance with Young Children* (Kersey and Masterson 2012). Professional development for in-service early childhood educators in social-emotional learning focuses first on responsive, nurturing, and attentive teacher-child interactions and relationships (Haslip, Allen-Handy, and Donaldson 2018) which closely aligns with the item cluster labelled 'spiritual responsiveness'. Fostering an environment of creativity, empathy, collaboration, cleanliness, and outdoor play (associated with 'promoting spiritual growth through the environment') also reflects core principles in early childhood education found in texts on the classroom environment (Curtis and Carter 2014; Masterson and Bohart 2019; Wien 2013). The practices related to 'promoting spiritual growth through the curriculum', however, are considered more advanced because of their integration focus. For example, experienced teachers are more likely to implement integrated projects, such as those described in *Young Investigators: The Project Approach in the Early Years* by Helm and Katz

Table 9. List of items, descriptions, and their Rasch difficulty in descending order of difficulty.

Cluster Title	Item Code	Item Description	Item Difficulty (Measure Score in Logits)
Promoting Transcendent Practices	2.21	Meditation into daily activities	2.18
	2.20	Yoga into daily activities	2.12
	4.34	Sharing spiritual/invisible realm experiences	1.90
Promoting Spiritual Connections	2.08	Talk about the meaningfulness of life	1.30
	3.27	Caretaking of plant and animal life	1.27
	2.22	Mindfulness into daily activities	0.89
	4.37	Interact and build relationship with nature	0.70
	4.33	Share wisdom with others	0.66
	3.28	Appreciation for aesthetics and beauty	0.53
	2.23	Relaxation/quiet times in curriculum	0.45
	2.13	Being a part of something greater	0.43
	3.31	Playing in nature	0.42
	2.09	Deeply enjoy surroundings	0.11
Promoting Spiritual Growth through the Curriculum	2.24	Expressive arts activities	−0.10
	2.19	Integrated curriculum projects	−0.12
	2.10	Deeply connect with others	−0.27
	4.40	A holistic understanding of child development	−0.48
Promoting Spiritual Growth through the Environment	3.30	Playing outside	−0.75
	3.29	Natural light	−0.77
	2.16	Think, move, and express emotion	−0.78
	4.38	Culture that encourages expression of virtues	−0.78
	2.18	Collaborative activities	−0.79
	2.12	Expressing joy activities	−0.87
	2.11	Engage in creativity	−0.94
	3.26	State of order and cleanliness	−0.97
	4.35	Demonstrate empathy for others	−1.00
	4.36	Show love for others	−1.15
Spiritual Responsiveness	2.15	Comment on spontaneous discoveries	−1.47
	2.14	Pay attention to questions/interests	−1.71

(2016). Learning how to facilitate a diverse array of expressive arts activities likewise requires experience, materials, and effort (Bea 2004) making endorsement of these practices more difficult than others. Trends also suggest a decline in emphasis on the arts in early childhood corresponding with an increased focus on early literacy and maths (Haslip and Gullo 2018).

The items in the cluster 'promoting spiritual connections' were more difficult to endorse than the responsiveness, environment, and curriculum-related item clusters. This is understandable because the easier clusters to endorse reflect common practices in developmentally appropriate practice (Copple et al. 2013) whereas the 'connections' items frequently reflect more nuanced and reflective views about spiritually nurturing education (e.g., children's sharing their wisdom; talking about the meaningfulness of life; being part of something greater) as found in the children's spirituality literature (Hart 2003; Hay and Nye 2006; Hyde 2008; Mata 2015; Mata-McMahon et al. 2018). Specific activities that require adjusting the schedule were hardest to endorse (e.g., meditation, yoga). This could also be related to the need for additional training in transcendent practices (Philibert 2017). Finally, allowing and even encouraging children to directly talk about their spiritual beliefs, experiences, and questions (e.g.,

about dreams, soul, afterlife, God) was also among the most difficult practices to endorse, likely given the societal belief in the United States that these topics are recognised as religious and therefore not included in curricula commonly utilised in secular educational settings.

Recommendations for instrument modification

Following psychometric analysis several modifications are recommended for the original instrument so educators or researchers can use it to measure educators' nurturing of children's spirituality. Specifically, the frequency rating scale analysis suggests collapsing the five-point scale to a four-point scale: 'Once a month or less', 'A few times a week', 'Daily', and 'Several times each day'. The scale should also include a 'Not Applicable' option for educators to report that a certain practice may not be appropriate or possible with their students. Including a 'Not Applicable' option does not hinder the psychometric analysis as missing data are not problematic for the Rasch model, like they are for CTT (Wright 1996). Additionally, the two reversed-scored items (2.17 and 4.39) should be removed as respondents were not responding to these items in a similar manner. These findings align with the anticipated poor psychometric performance of negatively worded items in instruments (Weems and Onwuegbuzie 2001). Revising the rating scale and removing the two problematic items are considered fundamental in order for the *ECE-SPC* to create meaningful measures. However, researchers may also make additional revisions to continue to improve the performance of the instrument.

The Rasch analysis identified several redundant items that measure approximately similar levels of the construct. For example, items 2.13, 2.23, 3.28, and 3.31, which range from 0.53 to 0.42 logits in difficulty, are not contributing unique measurements to the scale. As a result, if instrument developers seek to reduce the number of items, it may be beneficial to remove some of these items to create a shorter instrument. Removing redundant items would allow for new and more difficult/easier items to be added to ensure that there are items that exist beyond the range of respondents' ability level, which would contribute to better measurement of the construct (Boone, Staver, and Yale 2014; Wright and Masters 1982). However, even without these optional revisions, the existing instrument demonstrates excellent internal consistency. The decision to keep these items in the revised instrument was due to the fact that each of them reflects distinct practices early childhood educators could or could not be facilitating in their settings (e.g., providing quiet time, allocating time to play in nature, asking children to think about how they are part of something greater, etc.) and might be made aware of the possibility of facilitating them, if they are not doing so already, by being asked about them in the *ECE-SPC*. The final revised instrument is presented in Supplementary B.

Implications and future research

The findings from this study provide robust validity evidence to support the use of the *ECE-SPC* to measure educators' nurturing of children's spirituality. This validation study is the first of its kind regarding this type of instrument in the fields of early childhood education and early childhood spirituality. As a result, this instrument presents positive practical implications as it can be used by educators seeking to identify their own level of nurturing of children's spirituality (see Supplementary C). Administrators and professional development instructors may also use the instrument to evaluate educators' level of spiritual nurturing. Additionally, this study holds theoretical implications for future research.

Other researchers may use this instrument in their own investigation of nurturing of spirituality. Since this is the first instrument to be published via a peer-reviewed validation study, researchers can be confident in the measurement of this complex construct. Similarly, instrument developers may also take the recommendations in this study to continue to improve and adapt the instrument for other research and practical applications. For example, a larger national study would be beneficial to understand potential differences in psychometric performance based on educator demographics (e.g., race/ethnicity, gender) and administrator type more thoroughly. Researchers could also use the instrument to explore additional sources of validity evidence such as consequences of testing and additional convergent validity evidence with other data (e.g., Classroom Assessment Scoring System or CLASS data).

Limitations

This study was delimited by the focus on three of the five sources of validity evidence, which is consistent with other validation studies (Krupa, Bostic, and Shih 2020; Sondergeld, 2020). Response processes and consequences of testing sources may be explored in a separate qualitative study to address the remaining two sources of evidence. Specifically, as educators, administrators and researchers begin to use the instrument to measure changes in teacher practice over time or collect pre- and post-intervention measures after professional development activities, it is important to ensure the instrument is able to meet these needs through the study of consequences of testing validity evidence. The present study explores the claim that the *ECE-SPC* can measure educators' nurturing of children's spirituality at a single point in time. Claims that the instrument may be used to assess change over time or might influence teacher practices need to be explored in future studies. It is noteworthy to highlight this instrument has been designed for determining how and to what extent early childhood educators nurture spirituality in secular educational

settings and should not be used for other purposes such as measuring how nurturing spirituality impacts learning outcomes, or how children understand and express spirituality in the classroom.

Lastly, given that the participant sample reflects the larger early childhood workforce national population, with 77% being white females, the primary limitation of the study is the rather small sample size of non-white educators. Considering the few numbers of non-white and male educators, it was not possible to empirically explore potential differences in person measure scores. A more racial/ethnically and gender diverse sample would be beneficial to fully explore any potential relationships.

Conclusions

By exploring three of the five sources of validity evidence proposed by the *Standards* (AERA et al., 2014) (i.e., test content, internal structure, and relationship with other variables), through Rasch measurement theory (Rasch 1960, 1980), findings show that there was robust validity evidence for the *Early Childhood Educators' Spiritual Practices in the Classroom (ECE-SPC)* in measuring educators' nurturing of children's spirituality. Therefore, this validation study offers the early childhood education field the first instrument designed to measure how and to what extent educators are nurturing children's spirituality in secular educational settings. Furthermore, the study also contributes a Wright Map and calculator tool kit to help early childhood educators, school administrators, and centre directors interpret individual and aggregate scores after using the *ECE-SPC*. This tool kit will provide a method to better gauge what is being done, to what extent, and what could be incorporated to continue to support and nurture young children's spirituality in secular educational settings.

Disclosure statement

Notes on contributors

Jennifer Mata-McMahon, Ed.D., is an associate professor of early childhood education and director of the Sherman Center for Early Learning in Urban Communities at the University of Maryland, Baltimore County, in Baltimore, Maryland, U.S.A. Her research interests focus on children's spirituality and teacher preparation to nurture spirituality in secular educational settings. She is co-author of Ambiente en Acción (Environment in Action) (Unimet, 2006), author of Spiritual Experiences in Early Childhood Education (Routledge, 2015), co-editor of Spirituality: An Interdisciplinary View (Inter-Disciplinary Press, 2016), and The Bloomsbury Handbook of Culture and Identity from Early Childhood to Early Adulthood: Perceptions and

Implications (Bloomsbury, 2021). As well as the author of many book chapters and journal articles on children's spirituality.

Michael J. Haslip, Ph.D., is an early childhood education assistant professor at Drexel University, Philadelphia, Pennsylvania, U.S.A. His research interests include how to develop the character strengths of teachers and children, including spirituality. He also investigates how teachers use positive guidance strategies to improve children's social-emotional learning. He is the director of the McNichol Early Childhood Education Lab at Drexel University.

Lance Kruse, Ph.D., is the CEO of Invontics, LLC., a research company specializing in program evaluation, survey design, and psychometric and statistical analysis. His research interest includes instrument development, validation studies, and program evaluation within STEM education. As an interdisciplinary researcher, Dr. Kruse has conducted research in education, architecture, health care, and non-profit industries.

ORCID

Jennifer Mata-McMahon http://orcid.org/0000-0001-9222-0817
Michael Haslip http://orcid.org/0000-0001-9241-7994
Lance Kruse http://orcid.org/0000-0003-1706-2286

References

Allen, H. C., Ed. 2008. *Nurturing Children's Spirituality: Christian Perspectives and Best Practices*. Eugene, OR: Cascade Books.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (Eds.). 2014. Washington, DC: American Educational Research Association.

Bea, J. H. 2004. "Learning to Teach Visual Arts in an Early Childhood Classroom: The Teacher's Role as a Guide." *Early Childhood Education Journal* 31 (4): 247–254. https://doi.org/10.1023/B:ECEJ.0000024116.74292.56.

Berryman, J. W. 1991. *Godly Play: A way of religious eductaion*. San Francisco, CA: Harper San Francisco.

Bode, R., and B. Wright. 1999. "Rasch Measurement in Higher Education." In *Higher Education: Handbook of Theory and Research: Vol. XIV*, edited by J. Smart and W. Tierney, 287–316. Springer Netherlands. https://doi.org/10.1007/978-94-011-3955-7_7.

Bond, T. G., and C. M. Fox. 2015. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Third ed. United Kingdom: Routledge, Taylor and Francis Group.

Bone, J., J. Cullen, and J. Loveridge. 2007. "Everyday Spirituality: An Aspect of the Holistic Curriculum in Action." *Contemporary Issues in Early Childhood* 8 (4): 344–354. https://doi.org/10.2304/ciec.2007.8.4.344.

Bone, J., and A. Fenton. 2015. "Spirituality and Child Protection in Early Childhood Education: A Strength Approach." *International Journal of Children's Spirituality* 20 (2): 86–99. https://doi.org/10.1080/1364436X.2015.1030594.

Boone, W. J., J. R. Staver, and M. S. Yale. 2014. *Rasch Analysis in the Human Sciences*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4.

Bredekamp, S. 2016. *Effective Practices in Early Childhood Education: Building a Foundation*. Boston, MA: Pearson.

Briggs, D. C. 2004. "Comment: Making an Argument for Design Validity Before Interpretive Validity." *Measurement: Interdisciplinary Research & Perspective* 2 (3): 171–191.

Carpenter, W. A. 2003. "Jacob's Children and Ours: Richard of St. Victor's Curriculum of the Soul." *Educational Horizons* 82 (1): 44–54.

Chatterji, M. 2003. *Designing and Using Tools for Educational Assessment*. Boston, MA: Allyn and Bacon.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Copple, C., and S. Bredekamp, Eds. 2009. *Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth to Age 8*. 3rd ed. Washington, DC: National Association for the Education of Young Children.

Copple, C., S. Bredekamp, D. G. Koralek, and K. Charner, Eds. 2013. *Developmentally Appropriate Practice: Focus on Preschoolers*. Washington, DC: National Association for the Education of Young Children.

Cox, J. L. 2019. "Religion as the Transmission of an Authoritative Tradition: The Significance of Timothy Fitzgerald's Critique of Religious Studies for a Socially Embedded Definition of Religion." Implicit Religion 22 (3–4): 331–348. https://doi.org/10.1558/imre/40991

Curtis, D., and M. Carter. 2014. *Designs for Living and Learning: Transforming Early Childhood Environments*. St. Paul, MN: Redleaf Press.

Elkins, D. N. 1998. *Beyond Religion*. Wheaton, IL: The Theosophical Publishing House.

Erricker, J., C. Ota, and C. Erricker, Eds. 2001. *Spiritual Education: Cultural, Religious and Social Differences. New Perspectives for the 21st Century*. Brighton, UK: Sussex Academic Press.

Espinoza, B. D., J. R. J. Estep, and S. Morgenthaler, Eds. 2018. *Story, Formation, and Culture: From Theory to Practice in Ministry with Children*. Eugene, OR: Pickwick Publications.

Fisher, R. 1936. "Design of Experiments." *BMJ* 1 (3923): 554–554. https://doi.org/10.1136/bmj.1.3923.554-a.

Gandini, L., L. Hill, L. Cadwell, and C. Schwall, Eds. 2015. *In the Spirit of the Studio: Learning from the Atelier of Reggio Emilia*. 2nd ed. New York, NY: Teachers College Press.

Gill, S., and G. Thomson, Eds. 2016. *Redefining Religious Education: Spirituality for Human Flourishing*. New York, NY: Springer.

Goldstein, J., and J. K. Flake. 2016. "Towards a Framework for the Validation of Early Childhood Assessment Systems." *Educational Assessment, Evaluation and Accountability* 28 (3): 273–293. https://doi.org/10.1007/s11092-015-9231-8.

Gottlieb, A. 1998. "Do Infants Have Religion? The Spiritual Lives of Beng Babies." *American Anthropologist* 100 (1): 122–135. https://doi.org/10.1525/aa.1998.100.1.122 .

Gottlieb, A. 2006. "Non-Western Approaches to Spiritual Development Among Infants and Young Children: A Case Study from West Africa." In *The Handbook of Spiritual Development in Childhood and Adolescence*, edited by E. C. Roehlkepartian, P. E. King, L. Wagener, and P. L. Bensen, 150–162. SAGE Publications, Inc. https://doi.org/10.4135/9781412976657.n11.

Hambleton, R. K. 2000. "Emergence of Item Response Modeling in Instrument Development and Data Analysis." *Medical Care* 38 (9): II–60–65. https://doi.org/10.1097/00005650-200009002-00009.

Hart, T. 2003. *The Secret Spiritual World of Children*. Makawao, HI: Inner Ocean.

Haslip, M. J., A. Allen-Handy, and L. Donaldson. 2018. "How Urban Early Childhood Educators Used Positive Guidance Principles and Improved Teacher-Child Relationships: A Social-Emotional Learning Intervention Study." *Early Child Development & Care* 190 (7): 971–990. https://doi.org/10.1080/03004430.2018.1507027 .

Haslip, M. J., and D. F. Gullo. 2018. "The Changing Landscape of Early Childhood Education: Implications for Policy and Practice." *Early Childhood Education Journal* 46 (3): 249–264. https://doi.org/10.1007/s10643-017-0865-7.

Hay, D., and R. Nye. 2006. *The Spirit of the Child*. London, UK: Jessica Kingsley Publishers.

Helm, J. H., and L. G. Katz. 2016. *Young Investigators: The Project Approach in the Early Years*. New York, NY: Teachers College Press.

Hyde, B. 2008. *Children and Spirituality, Searching for Meaning and Connectedness*. London, UK: Jessica Kingsley Publishers.

Kane, M. T. 2013. "Validating the Interpretations and Uses of Test Scores." *Journal of Educational Measurement* 50 (1): 1–73. https://doi.org/10.1111/jedm.12000.

Kersey, K., and M. Masterson. 2012. *101 Principles for Positive Guidance with Young Children: Creating Responsive Teachers*. London, UK: Pearson Higher Ed.

Krupa, E., J. Bostic, and J. Shih. 2020. "Validation in Mathematics Education: An Introduction to *Quantitative Measures of Mathematical Knowledge: Researching Instruments and Perspectives*." *Quantitative Measures of Mathematical Knowledge: Researching Instruments and Perspectives* 1–13. Routledge. https://doi.org/10.4324/9780429486197-1.

La Paro, K. M., R. C. Pianta, and M. Stuhlman. 2004. "The Classroom Assessment Scoring System: Findings from the Prekindergarten Year." *The Elementary School Journal* 104 (5): 409–426. https://doi.org/10.1086/499760.

Larson, M. L., and R. J. Keeley, Eds. 2020. *Bridging Theory and Practice in Children's Spirituality. New Directions for Education, Ministry, and Discipleship*. Grand Rapids, MI: Zondervan Reflective.

Lawson, K. E. 2012. *Understanding Children's Spirituality: Theology, Research, and Practice*. Eugene, OR: Cascade Books.

Lawson, K. E., and S. May, Eds. 2019. *Children's Spirituality: Christian Perspectives, Research, and Applications*. 2nd ed. Eugene, OR: Cascade Books.

Linacre, J. 2002. "Optimizing Rating Scale Category Effectiveness." *Journal of Applied Measurement* 3 (1): 85–106.

Linacre, J. 2003. "PCA: Data Variance: Explained, Modeled and Empirical." *Rasch Measurement Transactions* 17 (3): 942–943.

Linacre, J. 2016. *Winsteps* (4.0.1) [Computer software]. Winsteps.com.

Linacre, J., and B. Wright. 2012. *A User's Guide to BIGSTEPS: Rasch-Model Computer Program*. San Diego, CA: Mesa Press.

Liu, X. 2010. *Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach*. Charlotte, NC: Information Age.

Mansour, N. 2009. "Science teachers' Beliefs and Practices: Issues, Implications, and Research Agenda." *International Journal of Environmental & Science Education* 4 (1): 25–48.

Masterson, M. L., and H. Bohart, Eds. 2019. *Serious Fun: How Guided Play Extends Children's Learning*. Washington, DC: National Association for the Education of Young Children.

Mata, J. 2012. "Nurturing Spirituality in Early Childhood Classrooms: The Teacher's View." In *Spirituality: Theory, Praxis and Pedagogy*, edited by M. Fowler, J. D. Martin, and J. L. Hochheimer, 237–248. BRILL. https://doi.org/10.1163/9781848880917_026.

Mata, J. 2015. *Spiritual Experiences in Early Childhood Education: Four Kindergarteners, One Classroom*. New York, NY: Routledge. https://doi.org/10.4324/9780203505175.

Mata-McMahon, J. 2017. "Spirituality and Humour: Making Connections for Early Childhood Education." *International Journal of Children's Spirituality* 22 (2): 170–178. https://doi.org/10.1080/1364436X.2017.1287681.

Mata-McMahon, J. 2018. "What Do Kindergarteners' Spiritual Experiences and Expressions Look Like in a Secular Classroom?" In *Story, Formation, and Culture: From Theory to Practice in Ministry with Children*, edited by B. Espinoza, J. R. Estep, and S. Morgenthaler, 235–252, Eugene, OR: Pickwick Publications.

Mata-McMahon, J. 2019. "Finding Connections Between Spirituality and Play for Early Childhood Education." *International Journal of Children's Spirituality* 24 (1): 44–57. https://doi.org/10.1080/1364436X.2019.1619528.

Mata-McMahon, J , Haslip, M , and Schein, D . 2018. "Early Childhood educators' Perceptions of Nurturing Spirituality in Secular Settings." *Early Child Development & Care* 189 (14): 2233–2251. https://doi.org/10.1080/03004430.2018.1445734 .

Mata-McMahon, J , Haslip, M , and Schein, D . 2020. "Connections, Virtues, and Meaning-Making: How Early Childhood Educators Describe Children's Spirituality." *Early Childhood Education Journal* 48 (5): 657–669. https://doi.org/10.1007/s10643-020-01026-8 .

Miller, J. P. 2000. *Education and the Soul: Toward a Spiritual Curriculum*. Albany, NY: State University of New York Press.

Miller, J. P. 2007. *The Holistic Curriculum*. 2nd ed. Toronto, Canada: OISE Press.

Miller, J. P. 2019. "Holistic Education: A Brief History." In *The International Handbook of Holistic Education*, edited by J. P. Miller, K. Nigh, J. B. Marni, B. Novak, and S. Crowell, 5–16. Routledge.

Miller, R. 1988. "Holistic Education: A Radical Perspective." *The Holistic Education Review* 1 (1): 2–3.

Miller, J. P., K. Nigh, J. B. Marni, B. Novak, and S. Crowell, Eds. 2019. *International Handbook of Holistic Education*. New York, NY: Routledge.

Minor, C. V., and B. Grant. 2014. "Promoting Spiritual Well-Being: A Quasi-Experimental Test of an Element of Hay and Nye's Theory of Children's Spirituality." *International Journal of Children's Spirituality* 19 (3–4): 213–227. https://doi.org/10.1080/1364436X.2014.960916.

Mislevy, R. 2004. "Toulmin and Beyond: Commentary on Michael Kane's "Certification Testing as an Illustration of Argument-Based Validation." *Measurement: Interdisciplinary Research & Perspective* 2 (3): 135–170.

Montessori, M. 1963. *The Secret of Childhood*. B. B. Carter, Trans. Bombay, India: Orient Longmans.

National Archives 2018. The Bill of Rights: A Transcription Accessed 13 January 2021. Retrieved from: https://www.archives.gov/founding-docs/bill-of-rights-transcript

National Workforce Registry Alliance n.d. Our Resources Accessed 13 January 2021. Retrieved from https://www.registryalliance.org/our-resources/

O'Connor, B. P., M. R. Crawford, and M. D. Holder. 2015. "An Item Response Theory Analysis of the Subjective Happiness Scale." *Social Indicators Research* 124 (1): 249–258. https://doi.org/10.1007/s11205-014-0773-9.

Of Timothy Fitzgerald's Critique of Religious Studies for a Socially Embedded Definition of Religion." *Implicit Religion* 22(3–4): 331–348. https://doi.org/10.1558/imre.40991 .

Ota, C., and C. Erricker, Eds. 2005. *Spiritual Education: Literacy, Empirical and Pedagogical Approaches*. Brighton, UK: Sussex Academic Press.

Padilla, J.-L., and I. Benítez. 2014. "Validity Evidence Based on Response Processes." *Psicothema* 26 (1): 136–144. https://doi.org/10.7334/psicothema2013.259.

Pajares, M. F. 1992. "Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct." *Review of Educational Research* 62 (3): 307–332. https://doi.org/10.3102/00346543062003307.

Parker, K., J. Horowitz, A. Brown, R. Fry, D. V. Cohn, and R. Igielnik 2018. *What Unites and Divides Urban, Suburban and Rural Communities*. Pew Research Center Retrieved 05 January 2021. https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2018/05/Pew-Research-Center-Community-Type-Full-Report-FINAL.pdf

Pellegrino, J., L. DiBello, and S. Goldman. 2016. "A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments." *Educational Psychologist* 51 (1): 59–81. https://doi.org/10.1080/00461520.2016.1145550.

Peterson, C., and M. E. Seligman. 2004. *Character Strengths and Virtues: A Handbook and Classification*. Vol. 1. London, UK: Oxford University Press.

Philibert, C. T. 2017. *Everyday SEL in Early Childhood: Integrating Social-Emotional Learning and Mindfulness into Your Classroom*. Routledge. https://doi.org/10.4324/9781315461052.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. København, Denmark: Danmarks Paedagogiske Instiut.

Rasch, G. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests*. Expanded ed. Chicago, IL: University of Chicago Press.

Ratcliff, D., Ed. 2004. *Children's Spirituality: Christian Perspectives, Research, and Applications*. Eugene, OR: Cascade Books.

Reise, S. P., and J. M. Henson. 2003. "A Discussion of Modern versus Traditional Psychometrics as Applied to Personality Assessment Scales." *Journal of Personality Assessment* 81 (2): 93–103. https://doi.org/10.1207/S15327752JPA8102_01.

Roehlkepartian, E. C. 2012. "It's the Way You Look at Things" How Young People Around the World Understand and Experience Spiritual Development." In *Understanding Children's Spirituality. Theology, Research, and Practice*, edited by K. E. Lawson, 152–172, Eugene, OR: Cascade Books.

Sagberg, S. 2015. *Holistic Religious Education - is It Possible? The Complex Web of Religion, Spirituality and Morality*. Munster, Germany: Waxmann.

Schein, D. 2012. *Early childhood educators' perceptions of spiritual development in young children: A social constructivist grounded theory study* () [Doctoral dissertation, Walden University]. ProQuest Dissertations Publishing Retrieved 15 June 2021.

Schein, D. 2017. *Inspiring Wonder, Awe, and Empathy: Spiritual Development for Children*. St. Paul, MN: Redleaf Press.

Schilling, S. 2004. "Conceptualizing That Validity Argument: An Alternative Approach." *Measurement: Interdisciplinary Research & Perspective* 2 (3): 178–182.

Semega, J. L., M. A. Kollar, A. Shrider, and F. Creamer, (2020). *Income and Poverty in the United States: 2019* (Current Population Reports). U.S. Department of Commerce. U.S. Census Bureau. U.S. Government Publishing Office Retrieved 05 January 2021. https://www.census.gov/content/dam/Census/library/publications/2020/demo/p60-270.pdf

Shaw, F. 1991. "Descriptive IRT Vs. Prescriptive Rasch." *Rasch Measurement Transactions* 5 (1): 131.

Sireci, S. 2013. "Agreeing on Validity Arguments: Validity Arguments." *Journal of Educational Measurement* 50 (1): 99–104. https://doi.org/10.1111/jedm.12005.

Smith, R. 1996. "A Comparison of Methods for Determining Dimensionality in Rasch Measurement." *Structural Equation Modeling: A Multidisciplinary Journal* 3 (1): 25–40. https://doi.org/10.1080/10705519609540027.

Smith, E., K. Conrad, K. Chang, and J. Piazza. 2002. "An Introduction to Rasch Measurement for Scale Development and Person Assessment." *Journal of Nursing Measurement* 10 (3): 189–206. https://doi.org/10.1891/jnum.10.3.189.52562.

Sondergeld, T. A. 2020. "Shifting sights on STEM education quantitative instrumentation development: The importance of moving validity evidence to the forefront rather than a footnote." School Science and Mathematics, 120(5), 259–261. https://doi.org/10.1111/ssm.12410

Sondergeld, T. A., and C. Johnson. 2014. "Using Rasch Measurement for the Development and Use of Affective Assessments in Science Education Research." *Science Education* 98 (4): 581–613. https://doi.org/10.1002/sce.21118.

Steiner, R. 1997. *The Essentials of Education*. Hudson, NY: Anthroposophic Press.

Thomson, M. M., and J. L. Nietfeld. 2016. "Belief Systems and Classroom Practices: Identified Typologies of Elementary School Teachers from the United States." *Journal of Educational Research* 109 (4): 360–374. https://doi.org/10.1080/00220671.2014.968912.

Tolbert, L. V., Ed. 2014. *Exploring and Engaging Spirituality for Today's Children. A Holistic Approach*. Eugene, OR: Wipf and Stock Publishers.

Tomlinson, J., E. S. Glenn, D. R. Paine, and S. J. Sandage. 2016. "What is the "Relational" in Relational Spirituality? A Review of Definitions and Research Directions." *Journal of Spirituality in Mental Health* 18 (1): 55–75. https://doi.org/10.1080/19349637.2015.1066736.

U.S. Bureau of Labor Statistics 2021, May 7. Labor Force Statistics from the Current Population Survey: Household Data Annual Averages Accessed 7 May 2021. https://www.bls.gov/cps/cpsaat11.htm

Waugh, R., and E. Chapman. 2005. "An Analysis of Dimensionality Using Factor Analysis (True-Score Theory) and Rasch Measurement: What is the Difference? Which Method is Better?" *Journal of Applied Measurement* 6 (1): 80–99.

Weems, G. H., and A. J. Onwuegbuzie. 2001. "The Impact of Midpoint Responses and Reverse Coding on Survey Data." *Measurement and Evaluation in Counseling and Development* 34 (3): 166–176. https://doi.org/10.1080/07481756.2002.12069033.

Wien, C. A. 2013. *The Power of Emergent Curriculum: Stories from Early Childhood Settings*. Washington, DC: National Association for the Education of Young Children.

Wright, B. 1996. "Comparing Rasch Measurement and Factor Analysis." *Structural Equation Modeling: A Multidisciplinary Journal* 3 (1): 3–24. https://doi.org/10.1080/10705519609540026.

Wright, B., and J. Linacre. 1994. "Reasonable mean-square fit values." *Rasch Measurement Transactions* 8 (3): 370.

Wright, B., and G. Masters. 1982. *Rating Scale Analysis*. San Diego, CA: Mesa Press.

Wright, B., and M. Stone. 1979. *Best Test Design*. San Diego, CA: MESA Press.

Wright, B., and M. Stone. 2004. *Making Measures*. Chicago, IL: Phaneron Press.

Yang, Y., P. He, and X. Liu. 2018. "Validation of an Instrument for Measuring students' Understanding of Interdisciplinary Science in Grades 4-8 Over Multiple Semesters: A Rasch Measurement Study." *International Journal of Science & Mathematics Education* 16 (4): 639–654. https://doi.org/10.1007/s10763-017-9805-7.

Yusoff, M. S. B. 2019. "ABC of Content Validation and Content Validity Index Calculation." *Education in Medicine Journal* 11 (2): 49–54. https://doi.org/10.21315/eimj2019.11.2.6.