Investigating the psychometric properties of the ACEI global guidelines assessment (GGA) in four countries.

By: Belinda J. Hardin, Doris Bergen, Hsuan-Fang Hung

Abstract:

The ACEI Global Guidelines Assessment (GGA) was developed to provide an international assessment tool that can be used by early childhood educators to develop, assess, and improve program quality worldwide. This pilot study was conducted in four countries to investigate the psychometric properties of the GGA within and across different countries. A total of 168 programs and 336 early care and education professionals participated in this study from communities in the People's Republic of China, Guatemala, Taiwan, and the United States. The results show strong internal consistency for each subscale and the total GGA as well as moderate interrater consistency for the five subscales. A comparison of item ratings and the qualitative evidence suggests moderately acceptable congruence between the ratings and evidence to support the ratings. Patterns of program practices were identified within and across the participating sties and countries that reflected both unique and common practices. These results suggest that the GGA has potential as a useful and effective tool both for understanding early childhood program quality within and across countries and as a means of helping practitioners to establish and/or improve the quality of their services.

Keywords: early childhood education | international education | program evaluation | education | international early childhood education

Article:

Introduction

As the number of children participating in early childhood care and education (ECCE) worldwide continues to grow, so does the need for policies, practices, and tools that support high quality experiences (United Nations Educational, Scientific, and Cultural Organization (UNESCO 2006a, 2010; United Nations Children's Fund [UNICEF] 2004). In 2004, for

example, almost 124 million children were enrolled in pre-primary education, 10.7 % more children than in 1999 (UNESCO 2006a). Demographic changes such as increased participation of women in the work force and rapid migration to urban areas without the benefit of extended family members are two examples of population changes that have created a need for out-of-home child care (UNESCO 2006a). At the same time, international organizations and national governments have advocated for more ECCE services based on research findings, such as the role of nurturing and stimulating environments on brain development in the early years and the positive impact high quality ECCE services can have on children's social, emotional, cognitive, language, and physical development (Belsky et al. 2007; Burchinal et al. 2010; National Scientific Council on the Developing Child 2007, 2010; NICHD Early Child Care Research Network 2000; Peisner-Feinberg et al. 2001; Shonkoff and Phillips 2000).

However, the quality of ECCE services varies worldwide and is often less than adequate, particularly in developing countries, but even in developed countries. Thus, while the expansion of ECCE programs is a positive development, there is a need for evidenced-based tools with global applicability that can help stakeholders make informed policy decisions to improve the quality of ECCE services for young children and ultimately have a positive effect on their school readiness and long-term outcomes.

Developing a program assessment tool with global applicability requires a balance between sensitivity to cultural differences and meaningful constructs that have validity across cultures. In other words, a global tool should reflect commonalities of ECCE services that reflect shared beliefs, values, and practices while, at the same time, flexibility in how these constructs are interpreted locally. A number of international organizations have attempted to identify specific elements of quality ECCE services (Meyers 2006; UNESCO 2006a). The existence of these frameworks and the common elements contained in them—the importance of environments, curriculum content, learning and teaching interactions that produce positive outcomes for young children—suggest that progress toward identifying a viable set of global indicators of program quality has begun to take shape. However, the need for a reliable and valid instrument designed by early childhood professionals from multiple countries and specifically for global use remains a challenge.

In recognition of this need, the Association for Childhood Education International (ACEI) developed the ACEI Global Guidelines Assessment (GGA), a self-assessment tool designed to assist early childhood professional in assessing and improving program quality (ACEI 2003, 2006; Barbour et al. 2004; Sandell et al. 2010). The GGA is based on the Global Guidelines for Education and Care in the 21st Century developed by the World Organization for Early Childhood Education (OMEP) and the Association for Childhood Education International (ACEI 1999) by more than 80 professionals representing 27 countries.

To date, the GGA has been translated into a dozen languages and is being used in many areas of the world. However, little systematic research has been conducted on this instrument. In

particular, the cultural validity of this instrument across these numerous settings were of concern. Thus, the ACEI Global Guidelines Task Force recommended that a pilot study be conducted in a small number of countries to begin to evaluate how reliable and culturally valid the results might be for countries with differing early childhood program characteristics. As a well-research instrument, the GGA has the potential to provide a common method of evaluating program quality that administrators, families, and policy makers can use to make informed decisions about policy changes and program practices. This article presents key findings of this pilot study, which was conducted across six sites in four countries—the People's Republic of China, Guatemala, Taiwan, and the United States—and revisions to the GGA resulting from this research.

Purpose of the Study

This pilot study was designed to begin answering how comparable and culturally valid the resulting information from the GGA is when used in a number of widely diverse settings. The selection of countries and sites was not random, however, but based on availability of ACEI researchers and colleagues in various parts of the world. Purposeful sampling procedures were used, however, to represent a broad range of variation in four ways, by: world region, country, program, and individual participant characteristics (Gall et al. 2007). The following questions guided this research:

1. What is the reliability and cultural validity of the GGA when examined across sites and countries as well as separately for each site?

2. What are the patterns of similarities and differences in program services by type of informant, site, country, and total sample?

3. When compared across countries and sites, do the results support the use of a single measure consistent with a global culture of early childhood program services while also acknowledging distinctions in program quality reflective of country and site differences?

Method

Geographic Areas and Characteristics

The four countries selected for the study included the People's Republic of China (China), Guatemala, Taiwan, and the United States, which represented three of the five regional areas designated by UNESCO: Asia and the Pacific, Europe and North America, and Latin America and the Caribbean (UNESCO 2006a). These countries were identified by ACEI Global Guidelines Task Force members who were from the particular country or had established relationships with early care and education professionals in the particular country. In Taiwan and the United States, two sites were recommended and included in the study because, in each case, one site was rural and the other urban. Only one site was identified in the other two countries.

A brief description of pre-primary education services is provided below for each country. The terms pre-primary, preschool, or kindergarten are used interchangeably to reflect the way the services are describe within the country context.

People's Republic of China Pre-primary education in China, which is called kindergarten, includes children who are 4–6 years of age. Approximately 36 % of pre-primary age children received ECCE services in 2003–2004 in about 117,900 kindergartens having a total enrollment of 20.8 million children (UNESCO 2006b). Teachers are required to be graduates of preschool education training programs, such as normal training schools, or take an exam locally supervised by a local educational authority (Corter et al. 2006; Wallet 2006). Kindergarten service formats range from large urban center to weekend classes, tutorial stations, child visit days, and home tutorial classes in rural areas (Corter et al. 2006; UNESCO 2006b, 2010; Wallet 2006). Nearly 20 % of the programs in this study were located in Dalian, China, a coastal city.

Guatemala Pre-primary (often called preschool or kindergarten) education in Guatemala officially includes children ages 4–6 years old (UNESCO 2006c), though younger children participate in child care services. In 2004, approximately 29 % of children ages 4–6 years old were enrolled in pre-primary education (UNESCO 2006c). Three types of ECCE services are offered, including programs operated by the Ministry of Education, the Secretaría of Social Well-Being-Presidency of the Republic, and private programs. The Ministry of Education provides pre-primary services either in public schools or through contracts with other education providers. Private programs are typically supported through fees paid by families or contracts with government agencies. The Secretaría's office provides services through the Centros de Atención Integral (Centers of Integral Attention in English, [CAI]) that include nutrition, education, full-day child care, health care, and parent involvement activities to working families. Ministry of Education teachers must be trained in pre-primary education in one of the 17 teacher training colleges, and child care staff are required to have two years of educational training. Participating programs in Guatemala comprised 18.4 % of the total sample, with all programs located in Guatemala City.

Taiwan Pre-primary education in Taiwan is provided in kindergartens and child care programs. Kindergartens are implemented by the Ministry of Education and enroll children from ages 4–6 years old, while child care programs accept children from one month to 6 years old. More than 201,800 children attended pre-primary services provided by both the public and private sectors in approximately 3,329 programs during 2006–2007 (Ministry of Education 2010a, b). Kindergarten teachers must be college graduates in education and childcare providers need an associate degree related to education. Two Taiwanese sites were included in this study, a 21-county area comprising 17.9 % of the sample and consisting of public kindergartens and primary school affiliated kindergartens mostly in urban areas, and another in a large rural, agriculture county, which included another 17.9 % of the sample.

United States In 2004, more than 2.1 million early childhood educators provided pre-primary services (called preschool for ages 3–4 years and kindergarten for 5-year-olds) in center, home, and school settings (National Association of Educators for Young Children 2010). Publically supported programs such as Head Start, funded by the federal government, provide comprehensive services to low-income families. Other public programs such as kindergartens and pre-kindergartens may be funded by a combination of federal and state funds. In 2004, there were more than 407,000 licensed/regulated ECCE programs in the US (National Association of Educators for Young Children 2010). Public school programs require teachers to have 4-year college degrees, while Head Start programs are still working toward this goal. Private program teachers must meet state regulations for education and training. These teachers may be required to have college bachelor's degrees, an associate's degree, or another type of certification. The US sample consisted of an Appalachian area in Ohio (13.1 % of the total sample), and the Triad Area of North Carolina (13.1 % of the sample).

Program and Participant Information

A total of 336 early childhood professionals (consisting of an administrator and teacher or designee for each program) from 168 programs in four countries across six sites participated in the study as indicated in Table 1.

Table 1

Total number of programs and participants by research site (N = 336)

Country	Total programs	Number of participants					
		Directors	Assistant directors	Teachers	Other	Missing data	Total participants
China	33 (20 %)	22 (14 %)	14 (66 %)	29 (21 %)	0 (0 %)	1 (25 %)	66 (20 %)
Guatemala	31 (18 %)	30 (19 %)	1 (5 %)	27 (20 %)	3 (25 %)	1 (25 %)	62 (18 %)
Taiwan 1	31 (18 %)	33 (20 %)	0 (0 %)	26 (19 %)	0 (0 %)	1 (25 %)	60 (18 %)
Taiwan 2	31 (18 %)	40 (24 %)	2 (10 %)	17 (12 %)	1 (8 %)	0 (0 %)	60 (18 %)
United States 1	22 (13 %)	18 (11 %)	0 (0 %)	20 (15 %)	5 (42 %)	1 (25 %)	44 (13 %)
United States 2	22 (13 %)	19 (12 %)	4 (19 %)	18 (13 %)	3 (25 %)	0 (0 %)	44 (13 %)
Total	168 (100 %)	162 (100 %)	21 (100 %)	137 (100 %)	12 (100 %)	4 (100 %)	336 (100 %)

Data for one participant were missing in these four sites, which is reflected in the total number of participants in each of these four sites

It should be emphasized that the following demographic information was reported by the program directors within their particular cultural context. The definitions of what constitutes a rural versus urban setting, for example, may vary from one country or site to the next. Likewise, it is important to point out that the way in which individual items on the GGA were rated reflects the local and societal contexts of the participants.

Most participating programs were private (63.1 %), another 26.8 % were public, and 10.1 % were reported as having funding from both private and public sources. Slightly more than half of the programs (53.9 %) were located in urban areas, 37.6 % in rural areas, and 8.5 % in a combination of rural and urban areas or reported as "other." Most of the families earned average incomes (69.4 %), and a smaller portion were poor (22.8 %), or wealthy (7.8 %). The majority of programs were open 10–12 months a year (87.9 %) and 9–12 h per day (83.8 %). There were ten or fewer classrooms in 85.2 % of the programs. Of the 127 programs that reported total enrollment, approximately two-thirds had fewer than 100 children (1–50 children at 26.0 % and

51–100 children at 40.9 %). Another 24.4 % had enrollments of 101–200 and the remaining programs (8.6 %) had enrollments ranging from 201 to 433 children. Most programs clustered children in single-age groups (69.1 %), others in multi-age groups (27.9 %), and a few used both multi- and single-age grouping (5.0 %). Services for both toddlers and preprimary children were provided by 38.0 % of programs, while 28.3 % served preprimary age children only. The remaining programs included the following age groups: infants through preprimary age (15.1 %); infants through primary age (13.3 %); infants through toddlers (2.4 %); and toddlers or preprimary through primary age (3 %). No program data were collected on the number of enrolled children with disabilities.

There were 335 individual participants who provided demographic information. All but 15 of the participants were female. The majority of the participants had a Bachelor's degree (41.6 %) or a 2-year degree or certificate (23.5 %). Only 2.8 % reported some graduate level training and 3.7 % of the participants (all from Guatemala) reported primary school as their highest level of education.

ACEI Global Guidelines Assessment

The 2006 edition of the GGA was used in this study. The first page of the GGA included space to record basic demographic information for the individual completing the assessment (name, gender, position in program, contact information). The GGA contained 88 items across five subscales: (a) Environment and Physical Space; (b) Curriculum Content and Pedagogy; (c) Early Childhood Educators and Caregivers; (d) Partnerships with Families and Communities; and (e) Young Children with Special Needs. Each subscale is further divided into topical subcategories. For example, Area 3: Early Childhood Educators and Caregivers include three subcategories: knowledge and performance, personal and professional characteristics, and moral/ethical dimensions. Each item is rated on a scale ranging from not available, inadequate, minimum, adequate, and good, to excellent, along with space to describe examples that support the selected rating as well as space for additional comments. Because there were no outside evaluator ratings, respondents were directed to provide examples that support their rating so that the congruence of the rating level and the evidence presented for the rating could be compared. See Sandell et al. (2010) for additional details about the history and content of the GGA.

Procedures

Data were collected from November 2007 to November 2008. Human subject procedures and country-specific permissions for the study, training of the on-site research coordinators (research site coordinators), and preparation of data forms were completed before the data collection

began. Also, research coordinators were trained and a system established for ongoing communication with them to ensure uniform data collection procedures.

Six research site coordinators implemented the study at the local level (three university faculty, two private program directors, and one doctoral student). Prior to beginning the study, research site coordinators participated in a 2-h conference call in which they were trained on confidentiality requirements, human subject requirements based on the Belmont Report (United States Department of Health, Education, and Welfare 1978), procedures for selecting programs, data collection procedures, and the GGA. In addition, the coordinators were required to contact the primary investigator at least weekly through email, phone calls, fax, or other methods of Internet communication. This process ensured that the research coordinators in each location were proceeding in congruence with the research requirements.

The coordinators' role was to initiate and supervise the on-site data collection activities within each country. Their responsibilities included recruiting local program participants, informing participants of data collection methods, disseminating and collecting all research materials, and assisting with questions or other information as needed. The coordinators identified potential programs through a combination of telephone calls, emails, and face-to-face recruitment visits. After each local program was recruited, the research site coordinator met with program directors to explain the purpose of the study and the procedures for data collection. Copies of the GGA and letters describing the study and requesting consent to participate in the study were shared and discussed with each program director in their native language. Written guidelines explaining the rating requirements were given to participants, including that: (a) raters must provide written examples to support their ratings; (b) two raters must independently conduct the GGA (a director/principal and a teacher familiar with the entire program); (c) both raters must complete the GGA during the same time period but without any discussion; and (d) ratings could not be discussed until the research copy of the GGA was put in a sealed envelope and returned to the research site coordinator. Each individual participant in the data collection received a certificate of participation from ACEI in appreciation of his or her participation. Completed assessments were mailed to the Principal Investigators for data entry and analysis.

A number of methods were used to address cultural validity issues. Translators and interpreters who were trained on the GGA procedures provided written and verbal communication as needed. They were native speakers with higher education degrees in either the language pertaining to the research site (e.g., English, Spanish, and Chinese) or the field of education. Also, local research coordinators and individual participants were asked to contact project staff for clarifications about the study procedures throughout the data collection process. It should be noted that GGA translations into the languages of participants were completed for ACEI prior to the study using the consensus method (Geisinger 1994) to ensure the cultural appropriateness of the translations.

Individual ratings on items and the evidence for the rating of each item were entered into a database. Ratings for each item were assigned a numerical value from 0 (not available), 1

(inadequate), 2 (minimum), 3 (adequate), 4 (good), to 5 (excellent) and entered into the database. Examples and comments serving as evidence for the ratings were translated into English for data analyses. Once all data were entered, two individuals verified the results for each item against the original protocol for both the quantitative and qualitative data. All errors were reconciled and corrected. Qualitative data were classified, coded, and organized into topical categories for each item by site. Coders examined the evidence for the rating and compared that evidence to the rating score using the following criteria: a code of 1 was indicated if there was little evidence to justify the rating score; a 2 if there was good but not sufficient evidence to justify the rating score; and a 3 if there was excellent evidence to justify the rating score. Coders were trained to rate the quality of evidence provided. Reliability of coder agreement across a random sample of responses showed a range of agreement between 98.6 and 99.1 %.

Results

The results of the pilot study are described below. The descriptive results are presented first followed by the reliability results, factor analyses findings, and the item rating/qualitative evidence results.

Descriptive Results

Item means were calculated for each of the 88 indicators and the total GGA. The item means, standard deviations, and ranges of the GGA results are shown in Table 2. The US1 ratings were typically higher than the other research sites and China ratings were somewhat lower on four of the five subscales than the other research sites. Tests of normality indicated that all but four items were within the normal range. Three of these four items were from Subscale 5 pertaining young children with special needs. The fourth item belonged to Subscale 4 and focuses on children and families visiting the program prior to enrollment, a practice that was reported as not typical. Types of items with high means for the total sample included those concerned with promoting good health practices with means ranging from 4.38 (0.54) in China to 4.86 (0.51) in US1; basic sanitation, nutrition, and potable water with means ranging from 4.30 (0.90) in Guatemala and 4.30 (0.65) in China to 4.71 (0.55) in US1; and providing environments that foster a sense of well-being for children that included means ranging from 4.21 (0.59) in China to 4.56 (0.59) in US1.

Table 2

Item means, standard deviations, and ranges for the program areas and total GGA by Country and for the total sample ($N = 335$)

GGA area	China (*n* = 66)			Guatemala (*n* = 62)			Taiwan1 (*n* = 60)			Taiwan2 (*n* = 59)			US1 (*n* = 44)			US2 (*n* = 44)		
	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
ENVPHY	3.85	0.80	2.71–4.38	3.95	1.02	2.89–4.67	4.01	0.95	2.80–4.70	4.11	0.85	3.09–4.69	4.34	0.71	2.77–4.86	4.18	0.80	2.84–4.65
CURPED	3.82	0.42	2.83–4.28	3.93	1.05	3.10–4.39	3.86	0.99	2.67–4.63	3.86	0.97	3.00–4.62	4.43	0.65	3.23–4.82	4.00	0.98	2.73–4.71
EDUCAR	4.08	0.66	3.78–4.38	4.19	0.92	3.85–4.50	4.14	0.74	3.83–4.55	4.18	0.82	3.91–4.53	4.54	0.56	4.23–4.75	4.30	0.77	4.05–4.56
PARCOM	3.71	0.91	2.86–4.38	3.57	1.38	1.92–4.38	3.62	1.18	1.50–4.75	3.85	1.02	2.07–4.71	4.44	0.68	3.89–4.80	4.00	0.97	3.07–4.61
SPCNED	3.09	1.21	2.19–4.29	2.51	1.75	1.00–4.33	3.45	1.35	2.17–4.50	3.48	1.46	2.19–4.46	4.39	0.87	3.33–4.84	3.41	1.46	1.40–4.57
TOT*GGA*	3.71	0.80	2.19–4.38	3.63	1.22	1.00–4.67	3.82	1.04	1.50–4.75	3.90	1.02	2.07–4.71	4.43	0.69	2.77–4.86	3.98	1.00	1.40–4.71

Many participants rated items in the last subscale pertaining to children with disabilities as "not available" or simply left these questions blank, reflecting the minimal provision of services for children with special needs. Differences in policy-related items were particularly evident. For example, item 81, "staff members are required to report plans for children with special needs to government agencies," was rated low by the Guatemala, US2, and China sites and somewhat higher by the remaining sites. In Guatemala, there are few services for children with disabilities, as only one diagnostic center with specialized services exists in Guatemala City. The US2 site was largely composed of private childcare programs. Even though these programs may have children with disabilities served through the local educational agencies (public schools), the programs themselves are not responsible for reporting information directly to the government. The findings in this subscale are similar to international monitoring reports examining services for children with disabilities in the Americas and Asia (Center for International Rehabilitation 2004, 2005) as well as worldwide reports (UNESCO 2010).

Reliability Results

Internal Consistency

Internal consistency estimates by subscale for the total GGA and the research sites were examined to determine how well the scale holds together as a single measure when used in different countries, and how well the items are measuring defined constructs (e.g., environment, physical space, curriculum, parent/community involvement, and special needs). Cronbach's alpha was used to calculate the internal consistency of each subscale by research site and for the total sample. The alpha coefficients for the total sample (N = 335) indicate very strong internal consistency for each subscale (0.89–0.92) and the total GGA (0.97). The alpha coefficients across the five subscales for each research site were strong as well (0.82–0.95) as was the coefficients for the total assessment by subscale (0.94–0.98). All alpha coefficients were significant at the p < .01 level.

Interrater reliability by participant type

Analyses were conducted to compare the ratings for each program between the two different participants. This analysis was conducted with 334 cases from 167 programs. Two participants from one Taiwan2 program were deleted due to incomplete data. Correlations using Pearson's r were computed between the two groups to examine the consistency in the ratings. Table 3 presents the means and standard deviations for each type of participant and the correlation coefficients across the two groups for each subscale and the total GGA. The resulting correlations indicate a moderate degree of consistency (0.46–0.70) for the five subscales and the total GGA (0.62) when completed by two different raters for the same program.

Table 3

Means, standard deviations, and correlations of GGA scores by type of participant ($N = 334$)

Program areas	Directors		Non-directors		r
	Mean	SD	Mean	SD	
ENVPHY	76.54	10.50	75.74	10.78	0.52
CURPED	65.66	11.84	66.36	11.05	0.50
EDUCAR	53.41	8.68	54.54	7.48	0.46
PARCOM	94.14	19.78	93.79	17.84	0.57
SPCNED	42.80	17.35	41.12	17.30	0.70
TOT*GGA*	332.55	56.96	331.54	53.79	0.62

For all correlations $p < .01$

To further understand factors influencing interrater correlations, additional analyses were conducted controlling for individual (highest education, number of years working at program, number of years in ECCE) and program demographic variables (program type, ages served). No significant differences were found when controlling for highest education level achieved and number of years working in early childhood care and education at the $p < .05$ level. No significant differences were found when controlling for number of years working in the current program, program type, and ages served at the $p < .01$ level.

Next, an additional analysis was conducted to examine interrater reliability for each research site. Three sites (China, Guatemala, and Taiwan1) had moderate to high correlations at the $p < .01$ level (0.41–0.80) and for each subscale and the total GGA (0.60–0.81). The US1 had good correlations for three subscales and the total GGA (0.60–0.74) at the $p < .01$ level. The Curriculum and Pedagogy subscale for the US2 site had a correlation of 0.61 at the $p < .01$ level also. However, this same subscale in the US1 site and three subscales and the total GGA for the US2 site had moderate correlations (0.48–0.51) at the $p < .05$ level. The Taiwan2 site correlations for all subscales and total GGA were not significant, as well as the Environment and Physical Space subscale for the US1 site and the Young Children with Special Needs subscale for US2.

Factor analysis

To understand the extent of interrelationships among the individual 88 items, a set of exploratory factor analyses was conducted with the total sample. The results for the total sample were compared to look at similarities and differences in explained variance. Principal component analysis of the total sample (N = 335) yielded 18 variables with eigenvalues greater than 1. Four distinct factors were revealed accounting for 31.5, 6.3, 3.8, and 3.2 % of the variance, respectively. A varimax rotation of the principal components indicated that the factor patterns were similar and therefore stable. Items with loadings of ±0.40 or higher were included in the final designation of each factor. Items that had double loadings on the four factors were only included in the factor in which it loaded at the ±0.40 or higher. The first factor represented general program quality and contained positive loadings on 85 of the 88 items across all five subscales. The second factor included positive loadings on 12 items, 10 of which were from the fifth subscale, Young Children with Special Needs. The remaining two loadings were from the fourth subscale and pertained to family and community partnerships, including joint decision-making and established partnerships in program planning, management, and evaluation. The third factor included three loadings from the fourth subscale concerning parent and community partnerships. One loading was positive and related to program policies that directly or through links with other community resources provide supports to families. The remaining two loadings were negative and pertained to whether connections between home and program are encouraged and maintained and if children and families could visit the program before starting to attend it regularly. The fourth factor represented environment and physical space and included three positive loadings focused on how well the indoor and outdoor environments supported children as constructors of learning.

Congruence of Raters' Scores in Relation to Evidence Provided

Because researchers were not able to do on-site verifications of respondents' ratings, an alternative method for verifying ratings was also piloted. This type of verification was designed to assess the degree of congruence between the rating scores and the evidence provided by the raters. The method was used to evaluate how much relevant evidence participants provided to justify their rating scores. The scores (1 = little evidence to justify the rating score, 2 = good but not sufficient evidence to justify the rating score, and 3 = excellent evidence to justify the rating score) were compared by research site and by position (directors vs. nondirectors) for the five subscales of the instrument. In general, the rating system worked well; however, the evidence provided by the participants reflected some cultural differences. For example, the item "The environment and physical space are free from hazards including unsafe equipment, pollution, and violence" was rated good by respondents from China, Guatemala, and US, but their evidence differed. The Chinese rater's evidence was "There are a few parking lots outside the school, it may cause some danger for children" while the Guatemalan rater indicated, "There is some violence in the community (by gangs)" and the US rater stated, "We plug covers and keep our chemicals out of reach." Similarly, raters from the three countries who gave a rating of excellent to the item, "There are opportunities for frequent and positive child–child and child–adult

interactions," gave this evidence: "Teachers organize different activities to have children share their happiness with others through games and discussion" (China); "Our most important achievement is that they enjoy and see their class as their second home" (Guatemala); and "Teacher and staff socialize with children all the time, fun music + silly stories" (US).

Table 4 shows the comparison by directors versus nondirectors and Table 5 shows the comparison by research site and position. For the most part, the scores were in the mid-range across programs and participants. However, there were significant differences ($p < .001$) among research sites in level of scores on all subscales: $F_{1,5} = 11.41$ for environment, 24.07 for curriculum, 17.79 for educators, 13.60 for parent/partners, and 10.43 for special needs. Tawain2 and US2 had lower congruence between the rating and evidence provided on all subscales. This was due less to the evidence being incongruent with the rating and more due to the fact that raters at these two sites provided either no evidence or minimal evidence to support the rating. That is, raters from these two sites gave high ratings to their sites on many of the dimensions but they did not provide evidence that supported the high ratings. Since this problem occurred in sites in two different countries, it did not appear to be a cultural validity issue.

Table 4

Means and standard deviations for congruence scores by type of participant ($N = 322$)

Program areas	Directors ($n = 156$)		Non-directors ($n = 166$)	
	Mean	SD	Mean	SD
ENVPHY	2.05	0.47	2.01	0.52
CURPED	1.95	0.46	1.90	0.52
EDUCAR	1.83	0.48	1.82	0.53
PARACOM	2.03	0.45	1.94	0.50
SPCNED	1.98	0.48	1.92	0.51
TOTGGA	1.97	0.47	1.92	0.52

Table 5
Means and standard deviations for congruence scores by type of participant and research site ($N = 335$)

Particip ant type	China			Guatemala			Taiwan1			Taiwan2			US1			US2		
	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*	M	SD	*n*
Non-director	2.07	0.50	42	1.95	0.30	31	2.32	0.30	24	1.58	0.34	19	2.19	0.63	25	1.84	0.67	25
Director	2.22	0.34	21	2.08	0.39	30	2.29	0.33	28	1.79	0.36	40	2.12	0.60	18	1.92	0.65	19
Total	2.12	0.46	63	2.01	0.35	61	2.30	0.31	52	1.72	0.36	59	2.16	0.61	43	1.87	0.65	44

For the research sites as a whole, there were few differences in the rating evidence provided by directors and non-directors. However, there was one position effect, with directors giving more evidence for partnerships, $F_{1,1} = 4.617$, $p < .05$. This may be due to directors having more information about the site's partnership activities. There was no research site by position effect, however, thus supporting the cultural validity across sites.

Discussion

In this pilot study, the ACEI Global Guidelines Assessment was piloted in four countries and across six sites to investigate its effectiveness as a measure that had cultural validity and program evaluation usefulness. Overall, the results indicated the GGA was a viable option for understanding program quality in these four countries, and that it has the potential to be useful worldwide. In regard to the three questions of interest, there were some useful findings and some questions still to be answered.

Question 1: What are the reliability and cultural validity of the GGA scores when examined across sites and countries as well as separately for each site?

The psychometric properties of the GGA examined in this study suggest that it holds up well as a single measurement instrument across the varied research sites and countries. It evidenced good internal consistency with strong alpha coefficients for the total sample by subscale and the total GGA, and this consistency was demonstrated as well as for each research site. Interrater correlations were moderately strong for the total sample. However, differences were found when interrater correlations were examined by research site. In particular, the Taiwan2 correlations were not significant. When the Taiwan2 site was excluded from the analysis, moderately strong correlations were found across the other five sites. It may be that Taiwan2's weak correlations were specific to that site rather than a problem with the instrument. However, additional investigations of interrater reliability across multiple research sites and countries need to be conducted to further examine this aspect of the GGA. Four factors were identified that explained how individual items performed and their relationship to each other for the total sample. These factors generally aligned with four of the five subscales of the GGA.

The investigation of the congruence of the ratings with the evidence provided across countries and respondents was partially successful. That is, requiring raters to give evidence for the rating by providing examples did enable the researchers to check how the rating level was supported by evidence. For those respondents who provided examples that supported their ratings, a more realistic assessment of how well they had rated their program, especially in relation to the ratings of respondents from other countries, was useful. As noted earlier, respondents from some of the countries rated their programs very high on most dimensions and, while some of them provided good evidence for their ratings, others provided minimal evidence to support these ratings. In some cases, high raters provided no evidence. In contrast, respondents who rated their programs lower generally provided good evidence for their responses. Thus, in future studies, it will be very important to stress to raters that they must provide evidence for every one of their ratings, whether the rating is high or low.

Some respondents commented that sections of the instrument had repetitive questions and that is why they did not continue to provide evidence for their ratings. A final analysis of these data

using the Rasch model was conducted in order to determine the structure for the revised GGA. The Rasch analysis compared the patterns of respondents' ratings on the various items in each subscale and provided Wright maps that presented "corrected" person measures (the number answering each item with same rating) and "corrected" item calibrations in Log Odds Units. When a number of items distribute at the same level, it is likely the items are measuring similar constructs and so can be removed without compromising the instrument. After this analysis, 12 items on the GGA were removed and another seven items were reworded to reduce redundancy. This third edition of the GGA was approved by the ACEI Global Guidelines Task Force in April 2011. It contains 76 items across the five program areas and it the version to be used in subsequent research studies. A larger international study that will include 18–20 countries across five world regions is now in process. This study will include a subsample of external observers in addition to the program raters.

Question 2: What are the patterns of similarities and differences in program services by type of informant, site, country, and total sample?

This study sought to understand how the GGA functions as a measure of program quality from multiple perspectives that reflected both common characteristics across countries and sites as well as individual differences. Meyers (2006) stresses the importance of examining program quality from a variety of perspectives, particularly in cross-cultural studies, in order to address issues of cultural validity. Overall, the US1 site had a pattern of higher ratings than the other sites for the overall sample and research sites for each subscale, perhaps due to availability of resources and policies as indicated by participant comments.

Specific item patterns on program quality issues were evident across sites and countries. First, for items focused on the degree to which children, families, or community representatives were engaged in self-evaluation, decision-making, or program planning and management, means were lower across all sites but US1. One explanation for this discrepancy in patterns might be a disconnect between the underlying philosophy represented by these items, (e.g., value of self-governance), and cultural practices accepted in many communities, which view these items as the responsibility of professionals and or authority figures as indicated in some participants' comments and other international studies (Meyers 2006; UNESCO 2006a).

Second, in the environment and physical space section, items pertaining to availability of outdoor play materials and children's level of interactions with outdoor materials were generally rated lower than other items. These results were not surprising since other program evaluation instruments often show low ratings for outdoor play environments and experiences, suggesting the GGA's sensitivity is similar to other measures of program quality on this issue (Halfon et al. 2009; Tietze et al. 1996). The reason for low ratings may be due to lack of outdoor equipment and materials, and/or a minimal emphasis on the value of outdoor play in the early childhood field generally.

Third, Subscale 5 ratings (special needs) were lower and more diverse than the other subscales across sites, with the exception of the US1 site and, to a lesser degree, the US2 site. The clustering of these items was confirmed in the factor analysis (second factor). Many participants identified limited resources and a lack of training and professional knowledge as major barriers. Though most participants supported the principle of inclusion, some felt this practice was not fully supported by their community.

Some differences in individual item means also provided insights about local cultural contexts. For example, in the first subscale Guatemala rated item 1, "the environment and physical space are free from hazards, including unsafe equipment, pollution, and violence" lower than the other sites. This appears to be an accurate response because of immediate dangers outside the walls of many of these programs, as well as close proximity to pollution (e.g., air pollution, trash, and debris). In China, item 37 was rated lower than all other sites. This item focused on teachers' knowledge and application of child development and pedagogical practices. According to international reports, though China has improved teacher training for ECCE professionals, their skills in these area continue to be a challenge (Corter et al. 2006; UNESCO 2006b). In sum, the results suggest that the GGA identified patterns of similarities and differences of program quality on multiple levels. However, some cultural validity issues remain in need of further exploration.

Question 3: When compared across countries and sites, do the results support the use of a single measure consistent with a global culture of early childhood program services while also acknowledging distinctions in program quality reflective of country and site differences?

In general, the GGA enabled the participant raters to find both excellent program quality areas and areas in which they would like to improve their programs. These areas varied across countries and sites. After the data were collected, staff members from some participating programs commented on how using the GGA affected their perceptions of the quality of their own programs. For example, program administrators in the Appalachia site indicated that they would continue to use the instrument at the programs they represented. In one Appalachian county, a meeting was held to discuss the GGA results and issues of concern flagged by the instrument were noted. In particular, the lack of consistency among services for children and families, inequity among programs in quality of resources for indoor and outdoor classrooms, and the need to access more volunteers and community resources were discussed and plans were made to prioritize budgets, distribute resources across programs, contract for equal access of resources, and schedule professional development.

In the Taiwan2 site, study participants stated that the GGA served as a reflective tool for them to use to conduct a self-evaluation. They noted that parents typically played a passive role in collaborating with school professionals, parent participation tended to be limited to attending school activities, and questions related to communicating with parents about child expectations were noted. This reflection brought a new perspective for teachers. One participant said, "We

need to keep close communication with parents about children's behaviors and academic performance…[and]… discuss further about what is our expectations towards children…."

Limitations

A number of limitations should be noted. First, because this was a pilot study of the GGA, a relatively small number of programs and individuals participated. Additional research with more programs and countries needs to be conducted to further examine the reliability and validity of the GGA scores. Implementation of the larger scale study mentioned earlier will provide data from a larger variety of programs to address this issue. This study will include data from approximately 750 programs, and 1,500 individuals to further understand how the GGA functions worldwide. The GGA that was revised in 2011 based on the results of this study will be used in the new research project. Second, an investigation of the concurrent (or criterion) validity needs to be conducted to examine the correspondence between individual scores on the GGA with scores on a similar instrument that has been used to judge program quality in early childhood settings (e.g., Early Childhood Environment Rating Scale-Revised) as well as test–retest reliability and interrater reliability with observers who are not staff members need to be conducted. All three of these studies will be included in the new international research currently underway. Third, raters may need to be more fully trained on the GGA prior to using it, particularly since the results are culturally bound and there are no set criteria for what constitutes an excellent versus good rating. Fourth, questions about the number of children with disabilities served needs to be added to the program information form to have a clearer understanding of how many programs are providing these services. Fifth, ultimately research on the relationship between quality programs according to GGA ratings and child outcomes needs to be conducted. Various types of child outcomes should be examined (e.g., developmental, academic, and social outcomes).

Conclusion

The GGA is grounded in the belief that interactions among and across a variety of settings contribute to children's development (Bronfenbrenner 1979; Bronfenbrenner and Morris 1998). Thus, knowledge gained from various levels of a system may influence and inform other levels. Information gained from classroom, program, country, and world levels can have a positive impact on the development of children in ECCE programs. The purpose of this study was to provide a first step toward establishing the reliability and validity of the GGA. There are many challenges in conducting international research, particularly in developing countries, but the need for quality measures is well documented. This pilot study suggests the GGA has potential toward helping early childhood educators develop and implement quality services. One of the most powerful aspects of the GGA is that it was developed by an international group of ECCE experts

with the intention of being used globally. Often an instrument developed by one country may not be equally applicable to others. The international approach used to develop the GGA makes it unique among the tools available to ECCE professionals worldwide. One purpose of this instrument is to help programs examine and improve the quality of their services in order for children to experience a nurturing, safe, and stimulating environment. The content of this instrument clearly emphasizes family participation as well as equity and accessibility of services for all children. The GGA is not designed for ranking programs or making decisions about merit pay. Rather, it is designed to assist ECCE staff interested in developing new services or making improvements in existing services.

Since 2003 when the GGA was first published on the ACEI website, it has provided program and policy guidance for staff in the countries where it has been used. The availability of such an instrument, which has already demonstrated potential for improving early childhood programs in a variety of countries, will be likely to have an important role in improving both early childhood practice and educational policy. However, further steps are being undertaken to confirm its reliability and cultural validity. Ultimately, a final step will then involve investigations of the GGA's effect on child outcomes in various cultures around the world.

References

Association for Childhood Education International. (2003). *ACEI Global Guidelines Assessment*. Washington, DC: Author.

Association for Childhood Education International. (2006). *ACEI global guidelines assessment* (2nd ed.). Olney, MD: Author.

Association for Childhood Education International & World Organization for Early Childhood Education. (1999). *Global Guidelines for Education and Care in the 21st Century Care in the 21st Century*. Washington, DC: Association for Childhood Education International.

Barbour, A., Boyer, W., Hardin, B., & Wortham, S. (2004). From principle to practice: Using the global guidelines to assess quality education and care. *Childhood Education, 80*, 327–331.

Belsky, J., Burchinal, M., McCartney, K., Vandell, D. L., Clarke-Stewart, K. A., & Owen, M. T. (2007). Are there long-term effects of early child care?*Child Development, 78*(2), 681–701.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In E. Damon (Ed.), *Handbook of child psychology* (pp. 993–1028). New York, NY: Wiley.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*, 166–176.

Center for International Rehabilitation. (2004). *International disability rights monitor: Regional report of the Americas 2004*. Chicago, IL: Author.

Center for International Rehabilitation. (2005). *International disability rights monitor: Regional report of Asia 2005*. Chicago, IL: Author.

Corter, C., Janmohammed, Z., Zhang, J., & Bertrand, J. (2006). *Selected issues concerning early childhood care and education in China.* Background paper for Education for All Global Monitoring Report 2007.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*, 304–312.

Halfon, N., Russ, S., Oberklaid, F., Bertrand, J., & Eisenstadt, N. (2009). *An international comparison of early childhood initiatives: From services to systems*. New York, NY: The Commonwealth Fund.

Meyers, R. G. (2006). *Background paper prepared for the Education for All Global Monitoring Report 2007 Strong foundations: Early childhood care and education–Quality in program of early childhood care and education (ECCE)*. Paris, France: UNESCO Publishing.

Ministry of Education, Republic of China (Taiwan). (2010a). *Education in Taiwan*. Retrieved from http://english.ntcpe.edu.tw/public/Attachment/011114485271.pdf.

Ministry of Education, Republic of China (Taiwan). (2010b). *Summary of education at all levels: FY 2006-07*. Retrieved from http://english.moe.gov.tw/ct.asp?xItem=946&ctNode=1184&mp=11.

National Association for the Education of Young Children. (2010). *Critical facts about the early childhood workforce*. Retrieved from http://www.naeyc.org/policy/advocacy/ECWorkforceFacts.

National Scientific Council on the Developing Child. (2007). *The timing and quality of early experiences combine to shape brain architecture: Working paper No. 5*. Retrieved from http://www.developingchild.harvard.edu.

National Scientific Council on the Developing Child. (2010). *Early experiences can alter gene expression and affect long-term development: Working paper No. 10*. Retrieved from http://www.developingchild.harvard.edu.

NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71*, 960–980.

Peisner-Feinberg, E. S., Burchinal, M., Clifford, R., Culkin, M., Howes, C., & Kagan, S. (2001). The relation of preschool child care quality to children′s cognitive and social development trajectories through second grade. *Child Development, 72*, 1534–1553.

Sandell, E. J., Hardin, B. J., & Wortham, S. C. (2010). Using ACEI's global guidelines assessment for improving early education. *Childhood Education, 86*, 157–160.

Shonkoff, J., & Phillips, (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Tietze, W., Cryer, D., Bairrāo, J., Palacios, J., & Wetzel, G. (1996). Comparisons of observed process quality in early child care and education programs in five countries. *Early Childhood Research Quarterly, 11*, 447–475.

United Nations Children's Fund. (2004). *The official source of the state of the world's children 2005*. New York, NY: Author.

United Nations Educational, Scientific and Cultural Organisation. (2006a). *Education for all monitoring report: Strong foundations early childhood care and education*. Paris, France: Author.

United Nations Educational, Scientific and Cultural Organisation. (2006b). *Strong foundations: Early childhood care and education: China early childhood care and education (ECCE) programmes*. Geneva: UNESCO International Bureau of Education.

United Nations Educational Scientific and Cultural Organisation. (2006c). *Country profile prepared for the education for all global monitoring report 2007: Guatemala early childhood care and education (ECCE) programmes*. Geneva, Switzerland: UNESCO International Bureau of Education (IBE).

United Nations Educational, Scientific and Cultural Organisation. (2010). *EFA global monitoring report: Reaching the marginalized*. Paris, France: Author.

United States Department of Health, Education, and Welfare. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Washington, DC: OPRR Reports.

Wallet, W. (2006). *Pre-primary teachers: A global analysis of several key education indicators. Background paper for EFA Global Monitoring Report 2007*. Paris, France: UNESCO Publishing.