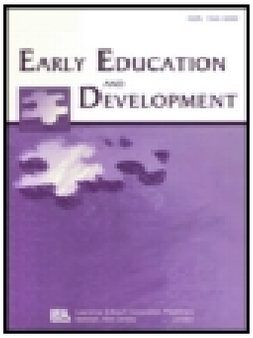

Routledge
Taylor & Francis Group

Early Education and Development

ISSN: 1040-9289 (Print) 1556-6935 (Online) Journal homepage: http://www.tandfonline.com/loi/heed20

Instrument Development and Validation of the Infant and Toddler Assessment for Quality Improvement

Michal Perlman, Ashley Brunsek, Anne Hepditch, Karen Gray & Olesya Falenchuck

To cite this article: Michal Perlman, Ashley Brunsek, Anne Hepditch, Karen Gray & Olesya Falenchuck (2016): Instrument Development and Validation of the Infant and Toddler Assessment for Quality Improvement, Early Education and Development, DOI: 10.1080/10409289.2016.1186468

To link to this article: http://dx.doi.org/10.1080/10409289.2016.1186468

Published online: 03 Jun 2016.

EARLY EDUCATION AND DEVELOPMENT
http://dx.doi.org/10.1080/10409289.2016.1186468

Routledge
Taylor & Francis Group

Instrument Development and Validation of the Infant and Toddler Assessment for Quality Improvement

Michal Perlman[a], Ashley Brunsek[a], Anne Hepditch[b], Karen Gray[b], and Olesya Falenchuck[a]

[a]Ontario Institute for Studies in Education, University of Toronto; [b]Children's Services, City of Toronto

ABSTRACT

Research Findings: There is a growing need for accurate and efficient measures of classroom quality in early childhood education and care (ECEC) settings. Observational measures are costly, as their administration generally takes 3–5 hr per classroom. This article outlines the process of development and preliminary concurrent validity testing of the Assessment for Quality Improvement (AQI), a new measure of global quality. The AQI is a classroom-level measure of structural and process quality. It consists of 24 items on a 5-point scale designed for use in ECEC infant and toddler classrooms. At between 60 and 90 min per room, the AQI is a relatively efficient measure. Item response theory modeling was used to ensure logical and coherent ordering of subitems. Exploratory factor analysis supported the use of the AQI total score and the Interactions section as a stand-alone measure. Correlations between the Infant and Toddler versions of the AQI and the Infant/Toddler Environment Rating Scale–Revised were moderate, providing preliminary support for the concurrent validity of both versions. *Practice or Policy*: Our results suggest that the AQI is a promising, efficient measure of global quality in infant and toddler ECEC environments. This may be especially relevant for Quality Rating and Improvement Systems, for which the observational component is a major cost driver.

Child Care Quality Matters

The early years are critical to children's brain development (Fisher, 2011), cognitive functioning (Landry & Smith, 2010; Whitebook, 2003), social and emotional functioning (Graves & Howes, 2011; McCartney, 1984; Squires, 2012), and the formation of attachment bonds (Read, 2014; Schore, 1994). Early experiences interact with genes to shape child well-being (Yoshikawa et al., 2013). Robust evidence suggests that children who attend higher quality early childhood education and care (ECEC) programs demonstrate better outcomes in these domains than do children in lower quality care (Burchinal, Roberts, Nabors, & Bryant, 1996; Clarke-Stewart, Vandell, Burchinal, O'Brien, & McCartney, 2002; Hamre & Pianta, 2001; Hestenes et al., 2015; Howes, 1988; Jeon & Buettner, 2015; National Institute of Child Health and Human Development Early Child Care Research Network, 2000; Peisner-Feinberg et al., 2001). High-quality care is also related to short-term improvements in children's early language, literacy, and mathematics skills (Camilli, Vargas, Ryan, & Barnett, 2010). ECEC quality is defined in terms of structural and process indicators. Structural quality indicators are related to the physical environment, organization of the space, and staff arrangements. Process quality indicators are related to the availability of learning opportunities for children as well as interactions between staff, between children, and between staff and children. Process quality is thought to impact children directly, whereas structural quality is thought to influence children

CONTACT Michal Perlman michal.perlman@utoronto.ca OISE/University of Toronto, 252 Bloor Street West, Toronto, ON M5S 1V6, Canada.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HEED.

indirectly through its impact on process quality (National Institute of Child Health and Human Development Early Child Care Research Network, 2002). Thus, ECEC programs are focusing improvement efforts on process quality in helping staff provide more cognitively stimulating opportunities and environments (Yoshikawa et al., 2013).

The Current Status of ECEC Quality

Although there is clear evidence of the impact of high-quality child care on child development, many children do not receive this level of care (Duncan, 2003; Goelman et al., 2006; National Institute of Child Health and Human Development Early Child Care Research Network, 2003; Peisner-Feinberg & Burchinal, 1997). Both longstanding and recent research has revealed that the average overall quality for ECEC programs is in the middle range on established measures of quality (Yoshikawa et al., 2013). High-quality infant and toddler care is especially difficult to find. A large-scale research study in North Carolina found that less than half (46%) of infants and toddlers in regulated care were in centers or homes with the highest quality ratings (4 or 5 stars under the state's 5-star rating system). In comparison, 58% of the state's preschoolers were in regulated care in centers or homes with the highest quality ratings (Carpenter, Martin, & Russell, 2005). The generally low quality of child care has led to calls to improve ECEC environments (National Early Childhood Accountability Task Force, 2007). Given the high proportion of Canadian children who receive some form of care from someone other than their parents (Bushnik, 2006), increasing attention is being paid to the quality of care that children receive and how it can be improved (Mitchell, 2005).

Child Care Quality Rating and Improvement Systems (QRIS)

One increasingly popular tool for quality improvement involves the development and implementation of quality rating systems. Quality rating systems are accountability systems that assess and communicate the level of quality in ECEC programs. Quality rating systems that explicitly include feedback and technical assistance as well as provide incentives to both motivate and support quality improvement are referred to as *Quality Rating and Improvement Systems* (QRIS; Tout et al., 2010). By defining quality standards, making program quality transparent, providing supports for improved inputs and processes, and disseminating information on quality ratings, QRIS are posited to promote quality of care, improve child outcomes, as well as provide a level of transparency that allows parents to make informed decisions about ECEC programs for their children (Zellman, Perlman, Le, & Setodji, 2008).

In the United States, QRIS have been implemented on a large scale at the state or county level (Tout et al., 2010). QRIS capture various indicators of quality such as staff–child ratios and staff training and education (Hestenes et al., 2015). The majority of QRIS also include an observational measure, most often one or more scales from the family of Environment Rating Scales (ERS) developed by Harms, Clifford, Cryer, and colleagues (Tout et al., 2010). These scales are designed to assess features of the learning environment such as the materials, activities, routines, provisions for health and safety, and interactions within the child care setting. More recently, as research continues to highlight the importance of process quality, some QRIS are placing a greater emphasis on the effectiveness of teacher–child interactions by using the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). The CLASS, an observational tool designed to quantitatively measure effective adult–child interactions, is being implemented in several state QRIS with the goal of capturing the level of cognitive stimulation provided by staff in ECEC environments (Tout et al., 2010). The CLASS and ERS are appealing because they are valid and reliable instruments. However, it is very costly to train, administer, and maintain reliability on these measures. For example, the ERS take between 3 and 5 hr per classroom to administer, whereas the CLASS takes approximately 3 hr per room. The cost of administering such measures annually, on a large scale, can be prohibitive. Some localities deal with the issue of cost by reducing the frequency of

assessments or by varying the method and frequency of monitoring specific standards. For example, Oklahoma monitors programs for licensing and overall QRIS compliance three times per year; however, ERS assessments are conducted only once every 3 years (Mitchell, 2005). Reducing the frequency of assessments can compromise the functioning of the QRIS. The City of Toronto has developed its own quality assessment tool with the goal of creating a more efficient, and in turn more sustainable, instrument for use in applied settings. The Assessment for Quality Improvement (AQI) is composed of three versions that are used in infant, toddler, and preschool classrooms. The purpose of this study was to refine and then test the psychometric properties and concurrent validity of the Infant and Toddler instruments and to determine their overall efficacy.

QRIS in the City of Toronto

Toronto's Children's Services division conducts detailed assessments of child care provider quality using the AQI. The AQI was initially developed by city staff over the course of the past three decades. It is a global, observational measure of classroom quality. A detailed description of the measure and how it is scored is described here and in more detail in the Appendix. The city employs a team of quality assurance analysts (QAAs) whose job consists of administering the AQI and who undergo interrater reliability testing every 3 months. AQI scores are used to determine whether the city will contract with a given program for the provision of subsidized child care spaces. Provincial legislation assigns the role of service system manager to Toronto Children's Services. Specifically, Children's Services manages, plans, and funds child care fee subsidies to families who need assistance. Approximately $318 million are provided annually by the Province of Ontario to manage the fee subsidy system. There are 25,116 fee subsidies currently available to help eligible families with the cost of child care. Approximately two thirds of licensed child care centers in Toronto provide care for children who receive a child care subsidy and therefore participate in the city's QRIS. As of January 2009, AQI scores have been posted online for use by parents and other stakeholders. This information is being used widely, with the relevant Web page on the City of Toronto website receiving more than 8,000 hits per month between September and December 2014. Results from AQI assessments are also used as a basis for quality improvement efforts. As parents' decisions about where to send their children are informed by these ratings and center reputations are influenced by them, the AQI is being used in high-stakes contexts.

The Infant and Toddler Versions of the AQI

The two versions of the AQI explored in this study measure quality in classrooms that serve infants (ages 0–18 months) and toddlers (ages 19–30 months). A substantial number of city staff and academics, as well as members of the early childhood community, have had input into developing the AQI (originally called the *Operating Criteria*).

The AQI is an observational measure consisting of multiple domains (33 in the initial version; see Table 1). Each domain is made up of varying numbers of subdomains (7–16 subdomains for the Infant measure and 7–18 domains for the Toddler measure) and is measured on a 5-point scale (1 or 2 = *does not meet expectations*, 3 = *meets expectations*, 4 or 5 = *exceeds expectations*). The AQI assesses various aspects of the classroom environment, including materials that are available for children within the program as well as interactions that occur between staff, between staff and children, and between children. Additional information about AQI scoring is provided in the Appendix.

In 2009 the psychometric properties of the city's Infant and Toddler AQI were tested using a sample of 235 infant classrooms from 199 child care centers and 386 toddler classrooms from 314 child care centers. Results of those analyses suggested that the psychometric properties of the two versions of the AQI were good, with Cronbach's alphas of .85 and .89 for the Infant and Toddler measures, respectively. However, validation data were not available at the time, and a clustering of scores at the high end of the 4-point scale was identified. The purpose of the current project was to

Table 1. Structure of the Initial Version of the Assessment for Quality Improvement Instrument.

Component	Domain	Domain Number
Structure of the Day	Daily and Visual Schedules	1
Activities and Experiences Planned	Program Plan	2
	Activities and Experiences	3
Physical Environment	Indoor Physical Environment	4
	Cloakroom Space and Storage	5
	Displays	6
Learning Areas	Sensory, Science, and Nature	7
	Art	8
	Books	9
	Language and Literacy	10
	Music and Accessories	11
	Physical Activities	12
	Blocks and Construction	13
	Cognitive and Manipulative	14
	Pretend Play	15
	Routine Care Practices	16
Physical Needs	Toileting and Diapering Routines	17
	Meals and/or Snack Time	18
	Equipment Required for Eating/Seating and Water/Refrigeration/Minor Food Preparation	19
	Cots or Cribs and Bedding	20
Health and Safety	Health and Safety	21
	Toys and Play Equipment Washing	22
	Children's Hand Washing or Sanitizing	23
	Staff Hand Washing or Sanitizing	24
	Transitions	25
	Attendance Verification	26
Interactions	Positive Atmosphere	27
	Supervision of Children	28
	Foster Children's Independence	29
	Supporting the Development of Self-Esteem	30
	Behavior Guidance	31
	Supporting the Development of Communication Skills	32
	Extending Children's Learning	33

develop and test the psychometric properties, reliability, and preliminary concurrent validity of the 5-point scale versions of the Infant and Toddler AQI.

The project was conducted in four phases. Phase 1 describes the activities and procedures that were involved in the development and refinement of the 5-point versions of the Infant and Toddler AQI. Phase 2 presents findings for a round of systematic field-testing with nearly 240 classrooms as well as the results of our item response theory (IRT) modeling that was used to revise the AQI. Phase 3 provides results of the analysis of the psychometric properties of the AQI based on operational data collected by the city in 2014, and finally Phase 4 presents concurrent validity data on a small sample of classrooms in which AQI scores were collected alongside other measures.

Phase 1: Measure Development

Objective

The objective of the first phase of this project was to review and refine the content of the 5-point Infant and Toddler versions of the AQI in order to make necessary revisions.

Method

Sample

The review of the Infant and Toddler versions of the AQI commenced with a literature review of current research and best practices related to quality in early learning programs. Next a total of 14 focus groups involving approximately 300 participants were conducted. Within Toronto Children's Services, several units manage and plan the different functions within the early learning sector. Recruitment for the focus groups targeted specific staff across each of these units to provide feedback on their particular area of expertise. These individuals represented the multiple stakeholders involved in the City of Toronto QRIS. These sessions were conducted between October and December 2012. Ten of the groups consisted of child care directors. For these groups, city staff sent out a flyer about the meeting, inviting all supervisors who were a part of the city's QRIS to participate in the focus group sessions. Demographic information was not collected in these sessions, but participants were almost exclusively women and reflected the highly ethnically diverse makeup of the ECEC workforce in Toronto. The recruitment process and broad demographics of the remaining groups are described next. They were broken down into the following stakeholder groups.

Academic experts in ECEC. Nine academic experts in the field of ECEC, child development, and program evaluation were invited by the principal investigator on the project to attend a focus group reviewing the measure in its entirety. All attendees were women. Four of the participants held an academic position at either a college or university. The remainder were directors of ECEC services at colleges and universities across the City of Toronto.

Aboriginal group. Six women representatives of the Aboriginal community were recruited through the Toronto Child Family Advisory Network. They were frontline staff, supervisors, and administrators.

French language operators. Approximately 20 women were recruited through the French Language Supervisor Network group. Administrators, supervisors, and frontline staff were all in attendance.

District consultants. A total of 21 district consultants (20 female, 1 male) were asked to review the instrument. District consultants manage a Children's Services portfolio within an assigned geographic area. They ensure that a range of accessible services for children and families are available and provide ongoing support to the service provider to ensure that issues related to financial and operational matters are dealt with. These individuals were recruited through an internal city process. The minimum education level for this group was a college diploma, and some attendees had a university degree. Because of scheduling constraints, feedback from the district consultants was collected vie e-mail rather than in a focus group format.

Four additional staff provided input on the entire Infant and Toddler versions of the AQI. Two staff were from the city's Special Services Unit. Children's Services, in partnership with community agencies, provides enhanced services to child care programs so that every child who needs extra support can actively participate in the program. Two staff from the corporate unit of the municipal government also reviewed the measure with an equity and accessibility perspective in mind.

Procedure

Attendees in the academic, Aboriginal, French, and district consultant focus groups were provided with copies of the measures in advance of their session and were asked to provide specific feedback regarding the content of each item. Participants in the child care director groups were provided with copies of the AQI during the sessions. Participants were asked to focus their analysis on the content of the items, their alignment with the domain, and the sequencing of the items within and across the categories of the measurement scale. The feedback from all of the focus group sessions was then

consolidated and reviewed by city staff. Changes to the instruments were incorporated based on this feedback. The revised version underwent additional field-testing by the QAAs, who then provided feedback on the feasibility of the changes. Revisions continued until no new revisions were suggested (i.e., saturation; Krueger & Casey, 2009) and city staff were satisfied. The final version was approved by the Children's Services Senior Management Team.

All review sessions were conducted over a 2-month period and were between 2 and 8 hr in duration. The supervisor network, Aboriginal, and academic expert groups were conducted in a semistructured manner with two facilitators present. The Special Services, corporate, and operator groups were conducted in a structured manner, with a facilitator asking a list of specific questions related to the content of the scales. As it was difficult to gather the district consultants, because of scheduling conflicts and location, they were e-mailed the AQI and given 1 month to review it in their own time and were asked to provide feedback via e-mail.

Results

After each focus group or review session, changes were made to the document based on participant feedback and this new version was then used for the next session. For example, several groups suggested that all subitems that pertained to the number of planned/documented activities in each learning area (e.g., "one planned/documented art experience," "one planned/documented sensory experience") be grouped under one domain (Domain 2: Program Plan) rather than under each of their respective domains (i.e., the Art domain or the Sensory domain). Participants thought that these subitems had to do with how and when ECEC staff set their programming and thus should fall under the Program Plan domain.

There were also several suggestions to modify the language that was used in the AQI to better reflect either the construct being measured or the age group being assessed. For example, the academic panel suggested that "Dramatic Play" be changed to "Pretend Play" to better reflect the skills and abilities of infants and toddlers. Furthermore, after we met with the corporate access and equity representatives and the Aboriginal focus group, it was suggested that the label "natural skin tones" (with regard to art materials) be changed to "diverse skin tones."

There was also some discussion about where subitems fell along the measurement continuum —that is, where subitems should be located depending on how easy or difficult they were to achieve. For example, the subitem "two or more science and nature activities are planned/ documented" was moved from "meets requirements" to "exceeds requirements," as focus group members thought that this was a more realistic standard given the resources available to ECEC staff. Similarly, "current documentation that observations of children are used in the development of activities is available" was moved from "meets expectations" to "exceeds expectations." Ordering of subitems in terms of increasing difficulty was further addressed empirically in Phase 2 of this study using IRT.

Phase 2: Field-Testing and IRT Analysis of the Infant and Toddler AQI Instruments

Objective

The objective of the second phase of this project was to field-test the Infant and Toddler versions of the AQI instrument in order to assess the measurement reliability and internal construct validity of this measure. The design of the instrument is based on the assumption that the items within each domain have a cumulative nature; therefore, to get a score on an item in a higher measurement category, a classroom would need to have scores on all preceding items. IRT modeling was used to test this assumption by empirically checking the properties and sequencing of the items within each domain and making modifications where required.

Method

Sample

The Infant and Toddler AQI instruments were field-tested for this phase of the project. The sample of classrooms used for this phase of the study was selected using a multistage stratified random sampling procedure to reflect the proportions of community- and school-based child care centers in Toronto. The sample consisted of 120 infant classrooms selected from 110 child care centers and 120 toddler classrooms from 119 child care centers, all having a contract to provide care for children receiving a child care subsidy with the City of Toronto. To ensure the time and resource efficiency of the data collection, we conducted the sampling in two stages: (a) Infant classrooms were selected, as some centers had toddler classrooms but no infant rooms; (b) toddler classrooms were selected from the centers from which infant classrooms were selected. This approach maximized the number of centers from which we could collect data from two classrooms. Classrooms were assessed between January and May 2013.

Procedure

Training of AQI assessors. Seven research assistants (RAs) were trained on the AQI and tested for reliability against a gold standard. Training involved 1 week of in-class review of the document and 1 week of field-testing in both infant and toddler rooms, followed by interrater reliability testing. Interrater reliability ranged from 96% to 100% agreement within one for infant rooms and from 88% to 100% agreement within one for toddler rooms, exceeding the 80% minimum required. In keeping with reliability standards used for other measures, such as the Infant/Toddler Environment Rating Scale–Revised (ITERS-R), scoring within 1 between the trainee and gold standard was considered a match. In order to minimize drift, we carried out a follow-up reliability test 3 months (about halfway) into the data collection period. All coders met reliability standards.

In centers in which three rooms were observed, one RA was responsible for administering the AQI in two rooms, going back and forth throughout the morning, as this is standard procedure during operational assessments, while a second RA collected data in the remaining classroom.

Data Analysis Plan

IRT is an advanced statistical technique that allows for the investigation of the psychometric properties of an instrument on a microlevel. It provides information that can be used to evaluate whether the items within each domain are placed in the right scoring category according to their level of difficulty (Hambleton, Swaminathan, & Rogers, 1991). IRT also allows for the detection of items that do not belong with the rest of the items in a specific domain. Separate IRT analyses were conducted within each of the 33 domains to identify the items that needed to be revised or removed from the instrument altogether. This was done separately for infant and toddler classrooms, for a total of 66 IRT analyses. The IRT analyses were conducted using XCalibre 42 software. The one-parameter logistic model was applied.

IRT modeling statistically estimates a parameter called *item difficulty*. This parameter quantifies how hard it is to attain a specific item within a domain. Item difficulty is an indicator that can take on negative and positive values on a continuum typically ranging from −3 to 3. More difficult items have larger positive values compared to easier items. Using item difficulty parameters for each item within a specific domain, it is possible to investigate whether the assumption of the cumulative nature of the items within a domain is satisfied. If this assumption is correct, then the items in higher scoring categories should have higher values for the item difficulty parameters. Item difficulty parameters can be used to modify the instrument by placing items within a domain sequentially according to their level of difficulty.

Results

The results for each of the 66 domains analyzed using IRT were outputted using the format displayed in Tables 2 and 3 and two charts (see Figure 1). Table 2 lists all items within each domain, their content definitions, and IRT difficulty parameters. IRT parameters could not be computed for items if they had no variance, zero valid responses, or low point biserial correlation. No variance means that either all or none of the observed classrooms attained this item. Such items do not add anything to the total domain score. However, such items should not be placed in Categories 3 or 4 and 5, as this may lead to score inflation. Zero valid responses means that the item is not applicable to either infant or toddler classrooms, as none of the classrooms in our sample exhibited this characteristic. Finally, a low point biserial correlation flags an item that is a candidate for removal from the scale because it has no relationship to other items within the domain.

Table 3 contains IRT difficulty parameters ordered by their magnitude. From this table we learn which sequence of the items within a domain would be statistically sound. To facilitate the processing of the information from the tables, we created the two charts that are displayed in Figure 1. The first chart (see Figure 1a) shows the distribution of item difficulties within each scoring category level (1 and 2, 3, and 4 and 5). Ideally, on this chart, we should see three boxes that do not overlap on the y-axis. This would mean that the current ordering of the items within a domain is correct. As can be seen from Figure 1a, the boxes for Level 1 and 2 and Level 3 overlap. This means that these scoring categories are not distinguishable by item difficulty. However, the

Table 2. IRT Item Difficulty Parameters.

Domain	Daily and Visual Schedules	IRT b parameter
Item 1/2a	Daily schedule is not posted	−0.782
Item 1/2b	Visual schedule is not accessible to the children	−0.447
Item 1/2c	Time is not planned for different activities/learning periods	Removed: no variance
Item 3a	Outdoor play	−0.762
Item 3b	Indoor play	−0.782
Item 3c	Child and staff initiated activities	−0.782
Item 3d	Children's physical needs	−0.782
Item 3e	Visual schedules is accessible to children	−0.447
Item 3f	Balance of structure and flexibility	−0.782
Item 4/5a	Alternate arrangements	1.770
Item 4/5b	Daily seasonally adjusted	1.301
Item 4/5c	Visual seasonally adjusted	0.944
Item 4/5d	Includes children in the program	0.051
Item 4/5e	Is referred to	1.498

Note. IRT = item response theory.

Table 3. Domain Structure by Increasing Value of IRT Item Difficulty Parameters

IRT Recommended Structure	
Item 1/2c	Removed: no variance
Item 1/2a	−0.782
Item 3b	−0.782
Item 3c	−0.782
Item 3d	−0.782
Item 3f	−0.782
Item 3a	−0.762
Item 1/2b	−0.447
Item 3e	−0.447
Item 4/5d	0.051
Item 4/5c	0.944
Item 4/5b	1.301
Item 4/5e	1.498
Item 4/5a	1.770

Note. IRT = item response theory.

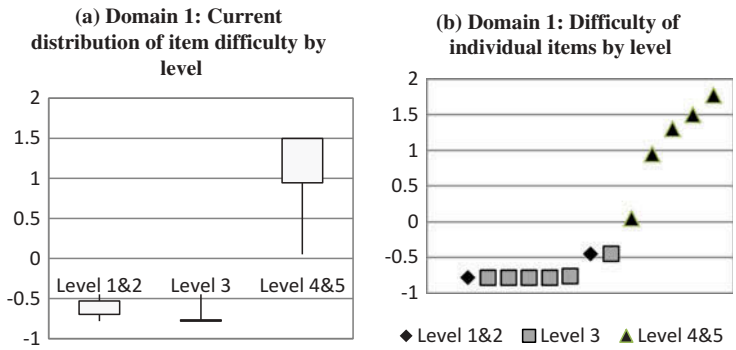

Figure 1. Distribution of item difficulty (a) by level and (b) by item in Domain 1.

box for Level 4 and 5 is located much higher on the graph, which confirms that the Level 4 and 5 items are ordered correctly.

Finally, the second chart (see Figure 1b) illustrates the suggested sequencing of items within a domain. Ideally, we want to see an upward trend of diamonds, followed by squares and then triangles. A mix of these symbols, as we see in Figure 1b with a Level 1 and 2 item being harder to achieve than most Level 3 items, suggests that a reordering of the items across the scoring categories is required.

Reordering of the items within the domains should be guided by statistical evidence as well as conceptual meaning of the quality indicators and policy implications. Therefore, the IRT analyses outputs were reviewed by the research team with input from the Toronto Child Care Service staff. Based on this review the following changes were made within and across the domains of the AQI instruments:

(1) Some items showing no variability (all classrooms had the same attribute) were removed.
(2) Most items showing low point biserial correlations were removed.
(3) Some items were moved across the scoring categories to reflect their actual difficulty level.
(4) Some domains were combined.

After the revisions, a new set of IRT analyses were performed and reviewed to examine whether the properties of each domain improved. This was repeated until the items within each domain of the Infant and Toddler instruments were structured based on item difficulty in increasing order, and content-based thresholds for assigning scoring levels from 1 to 5 were established for each domain.

Once the IRT analyses were completed, the instruments were reviewed by the quality assurance team at the City of Toronto and were approved by the Children's Services Senior Management Team. The final revised versions of the Infant and Toddler AQI instruments consisted of 24 domains each.

Phase 3: Psychometric Analysis of Operational Data (Structural Validity and Internal Consistency)

Objective

The objective of the third phase of this project was to test the psychometric properties of the revised Infant and Toddler versions of the AQI instrument to explore the structural validity and internal consistency of each measure. This was done using a convenience population sample—the operational data collected routinely by the city as a part of its QRIS. This sample consisted of all classrooms that served infants and toddlers and that participated in the city's QRIS.

Method

Sample

The sample used to test the psychometric properties of the final revised version of the Infant and Toddler AQI instruments consisted of 251 infant classrooms and 464 toddler classrooms from community- and school-based child care centers that had contracts with the City of Toronto. These made up 98% and 94% of all infant and toddler rooms, respectively, that were part of the city's QRIS. A small number of classrooms (four infant and 32 toddler rooms) were not included because they were assessed past the project end date.

AQI Infant and Toddler Instruments

In this phase the revised versions of the Infant and Toddler AQI instruments that had resulted from Phase 2 of this study were tested.

Procedure

In keeping with the City of Toronto's QRIS protocol, the city's QAAs administered all versions of the AQI after meeting a minimum standard of 80% exact match agreement. Re-reliability testing was conducted three times over the course of the year for each QAA. For this phase of the project, data collection was completed by nine QAAs. All QAAs were trained and reliable. To reduce bias, centers were randomly assigned to QAAs, ensuring that QAAs did not visit the same center 2 years in a row and that they visited centers across the city. QAAs collected data between May and December 2013.

Data Analysis Plan

The data analyses in this phase of the study included exploratory factor analysis and analysis of internal consistency of the measure. Exploratory factor analysis is a statistical technique that captures the underlying (latent) structure of a scale/instrument. Specifically, exploratory factor analysis allowed us to determine whether the domains of the Infant/Toddler AQI instruments form several subscales or whether one scale based on all 24 domains should be considered. Factor analysis is an important part of any validation study, as it provides evidence for computation of the subscale or scale scores based on a group of items (Izquierdo, Olea, & Abad, 2014).

Factor analysis uses correlation coefficients between individual items. Correlation coefficients of different types should be used depending on the scoring system of individual items. For example, for items that are continuous in nature (e.g., a person's age or income in dollars), Pearson correlations are used. For categorical (e.g., yes/no, pass/fail) items, tetrachoric correlations are used; for items scored on a Likert scale (e.g., 4-, 5-, or 7-point scales), polychoric correlations are appropriate. Because the AQI items are scored on a 5-point scale, factor analyses for this study were based on polychoric correlations.

Factor analysis involves a lot of decision making from the analyst. One of the decisions is about the number of subscales that should be considered for the instrument. An indicator used in making these decisions is called an *eigenvalue*. The number of eigenvalues produced in factor analysis is equal to the number of items in the scale. Only eigenvalues that are greater than 1 are considered to be indicative of potential item clusters in the solution. The goal of the analyst in factor analyses is to find a balanced solution that explains a sufficient amount of variability in the data and has good theoretical interpretability.

Once the analyst decides on the number of subscales that are present in the instrument or decides that the instrument represents one scale based on all of the items, Cronbach's alpha coefficients are computed for each group of items forming a subscale or for the full instrument. Cronbach's alpha is the index of the internal consistency of the scale. Cronbach's alpha values greater than .70 indicate that the items hang together well (i.e., they have a strong relationship with one another) and that a

reliable total score can be computed from these items. Cronbach's alpha and corrected item–total correlations were used to examine the internal consistency of the AQI instrument.

An additional issue that needs to be investigated is the validity of the reporting practices. Operational use of the AQI instrument involves evaluation of individual classrooms within child care centers. However, these scores are ultimately averaged within each program (i.e., for each center, all Toddler AQI classroom scores are averaged to create a single program score). This averaged score is then posted online and is used for other decision-making purposes. Given this procedure, it is important to investigate whether this aggregation leads to valid average scores for the quality of child care centers that accurately reflects scores for the individual classrooms within the centers. The average Infant/Toddler AQI score of the center would be legitimate if the scores of individual classrooms were reasonably consistent. However, if the difference between the scores of individual classrooms is too large, then the average score of the center would not reflect its actual quality, as it would not represent the actual quality of any classrooms within the center.

To investigate the variability of the Infant and Toddler AQI scores between classrooms within the centers, we used the subsample of centers in which more than one classroom was evaluated. Of the 251 child care centers that had infant rooms, 179 (71.3%) were excluded from these analyses as they had only one infant room. The remaining 72 centers had either two (23.9%) or three (4.8%) infant rooms. Therefore, a subsample of 72 centers with multiple infant rooms was used in this analysis.

Similarly, the operational data consisted of 464 centers with toddler classrooms. Of these centers, 54.1% had only one toddler room and were excluded from these analyses. The remaining centers had either two (38.4%), three (5.8%), or four (1.7%) toddler rooms. Therefore, the analyses of the Toddler AQI were based on a subsample of 213 centers.

Results

Score distributions on each domain of the Infant and Toddler AQI measures are reported in Tables 4 and 5. As can be seen from these tables, these distributions varied across the domains in both types of classrooms. For example, most classrooms attained high scores in the Program Plan and Art and Sensory domains (see Tables 4 and 5). The mean scores across all domains were above 3 in both types of classrooms, indicating that most centers tended to meet quality expectations. Note, however, that almost half of the classrooms did not meet expectations in the Learning Experiences and Staff and Children Hand Hygiene domains.

Each domain in the AQI measure is scored on an ordinal 5-point scale. Therefore, factor analyses were based on polychoric correlations between the domains. Factor analyses were performed using a principal axis factoring extraction method and promax oblique rotation. The first attempt to factor analyze the AQI measure was conducted with the request to extract all factors with eigenvalues greater than 1. Three factors were extracted in both the Infant and Toddler versions of the measure. In the Infant measure the three-factor solution accounted for 82.4% of the variance, and in the Toddler measure a similar solution accounted for 94.0% of the variance. Investigation of the factor structure in each of the versions showed that the domains related to staff–child interactions loaded together on the same factor, and most of the rest of the domains loaded onto the second factor. The third factor had high loadings for a few domains related to routines, health, and safety. Specifically, in the Infant version of the measure Domains 13 (Diapering Routines), 16 (Health and Safety and Toys and Play Equipment Washing), and 17 (Staff and Children Hand Hygiene) loaded onto the third factor, whereas in the Toddler version only Domains 13 and 17 loaded onto the same factor. The correlations between the factors were moderately strong in both versions of the measure: between .32 and .55 in the Infant AQI and between .43 and .55 in the Toddler AQI. The results of the three-factor solutions suggest that Interactions and Routines subscale scores can be computed for the AQI measure if needed. However, the pattern of correlations between the factors also suggests that a unidimensional structure might be considered. Therefore, the second round of factor analyses was conducted by requesting to extract a single factor. A one-factor solution accounted for

Table 4. Means, Standard Deviations, and Frequencies of Scores for Each Domain of the Infant Version of the Assessment for Quality Improvement (N = 251)

		Percentage of Classrooms						
		Does Not Meet Expectations		Meets Expectations	Exceeds Expectations			
No.	Domain	1	2	3	4	5	M	SD
1	Daily and Visual Schedules	0.0	17.5	43.8	25.5	13.1	3.34	0.92
2	Program Plan	0.0	10.4	5.6	20.7	63.3	4.37	0.98
3	Learning Experiences	0.0	42.2	3.2	28.3	26.3	3.39	1.27
4	Indoor Physical Environment	0.0	33.5	7.2	37.8	21.5	3.47	1.16
5	Displays	0.0	4.0	16.7	51.0	28.3	4.04	0.78
6	Art and Sensory	0.0	2.8	18.7	17.5	61.0	4.37	0.88
7	Books, Language, and Literacy	0.0	14.7	23.1	33.1	29.1	3.76	1.03
8	Music and Accessories	0.0	2.8	7.6	19.5	70.1	4.57	0.75
9	Physical Play Learning Experiences	0.0	1.6	11.2	26.7	60.6	4.46	0.75
10	Cognitive and Manipulative and Science and Nature	0.0	3.6	19.5	25.9	51.0	4.24	0.89
11	Blocks and Construction and Pretend Play	0.0	5.6	35.9	26.7	31.9	3.85	0.94
12	Routine Care Practices	0.4	28.3	4.0	10.8	56.6	3.95	1.33
13	Diapering Routines	0.0	39.4	15.9	28.3	16.3	3.22	1.14
14	Meals and/or Snack Time	0.0	45.8	1.6	16.7	35.9	3.43	1.37
15	Cribs and Bedding	0.0	3.2	0.0	10.4	86.5	4.80	0.59
16	Health and Safety and Toys and Play Equipment Washing	0.0	33.1	4.4	22.7	39.8	3.69	1.30
17	Staff and Children's Hand Hygiene	0.0	43.0	2.0	15.9	39.0	3.51	1.38
18	Transitions and Attendance Verification	0.0	29.9	6.8	23.9	39.4	3.73	1.26
19	Positive Atmosphere	0.0	0.0	12.4	29.1	58.6	4.46	0.71
20	Supervision of Children	2.4	6.0	9.2	17.9	64.5	4.36	1.03
21	Foster Children's Independence	0.0	4.4	9.2	17.9	68.5	4.51	0.84
22	Supporting the Development of Self-Esteem	0.4	8.8	38.2	29.9	22.7	3.66	0.94
23	Behavior Guidance	0.0	.8	6.4	14.3	78.5	4.71	0.62
24	Supporting Communication and Extending Children's Learning	0.0	.4	21.9	20.7	57.0	4.34	0.83

50.9% of the variance in the Infant AQI and 66% of the variance in the Toddler AQI. The factor loadings for individual domains varied between .28 and .67 in the Infant measure, with only two domains having factor loadings below .4 (Daily and Visual Schedules, and Physical Play Learning Experiences). In the Toddler version of the instrument, the factor loadings for individual domains varied between .25 and .66, with only two domains having factors loading below .4 (Cribs and Bedding, and Health and Safety and Toys and Play Equipment Washing). The one-factor solution was preferred, as it suggests the possibility of representing classroom quality with a single score without losing much information in the data.

To investigate the internal consistency of the AQI, we computed Cronbach's alpha coefficients. For the Infant version of the AQI Cronbach's alpha was .84, with item–total correlations for individual domains varying between .20 and .56. For the Toddler version of the instrument Cronbach's alpha was .86, with item–total correlations varying between .18 and .55.

Cronbach's alphas for a subset of domains reflecting the quality of teacher–child interactions (Domains 19–24) in the Infant and Toddler versions of the AQI were .80 and .82, respectively. Cronbach's alphas were not computed for the subset of domains that formed a third factor in the three-factor solution, as it is not recommended to create a scale from fewer than four items (Field, 2009).

Multilevel linear mixed models were fit to these data to explore the variance distribution within and between the centers. The amount of variance in the Infant AQI between centers was significant, Wald $Z(1) = 3.42$, $p < .01$, with about 73% of variability in the Infant AQI scores observed between the centers. Correspondingly, only about 27% of variance occurred between the classrooms within centers. Similarly, the amount of variance in the Toddler AQI between centers was significant, Wald $Z(1) = 5.79$, $p < .01$, with about 68% of variability in the Toddler

Table 5. Means, Standard Deviations, and Frequencies of Scores for Each Domain of the Toddler Version of the Assessment for Quality Improvement (N = 464)

		Percentage of Classrooms						
		Does Not Meet Expectations		Meets Expectations	Exceeds Expectations			
No.	Domain	1	2	3	4	5	M	SD
1	Daily and Visual Schedules	0.2	5.2	29.5	54.1	11.0	3.70	0.74
2	Program Plan	0.2	9.5	4.1	26.5	59.7	4.36	0.95
3	Learning Experiences	0.2	48.7	2.8	19.6	28.7	3.28	1.33
4	Indoor Physical Environment	0.2	7.3	18.8	46.1	27.6	3.94	0.88
5	Displays	0.0	1.1	10.8	47.4	40.7	4.28	0.69
6	Art and Sensory	0.0	4.5	2.2	14.2	79.1	4.68	0.73
7	Books, Language, and Literacy	0.0	13.6	15.1	49.8	21.6	3.79	0.93
8	Music and Accessories	0.0	3.9	6.3	22.0	67.9	4.54	0.78
9	Physical Play Learning Experiences	0.0	3.2	15.7	40.9	40.1	4.18	0.81
10	Cognitive and Manipulative and Science and Nature	0.0	1.1	1.3	33.0	64.7	4.61	0.57
11	Blocks and Construction and Pretend Play	0.0	10.3	3.4	29.5	56.7	4.33	0.95
12	Routine Care Practices	2.2	25.4	11.4	16.4	44.6	3.76	1.31
13	Diapering Routines	0.0	47.0	5.0	17.5	30.6	3.32	1.33
14	Meals and/or Snack Time	0.0	20.9	21.6	33.4	24.1	3.61	1.07
15	Cribs and Bedding	0.0	1.7	1.7	16.7	79.9	4.75	0.57
16	Health and Safety and Toys and Play Equipment Washing	0.0	23.1	5.4	24.1	47.4	3.96	1.20
17	Staff and Children's Hand Hygiene	0.2	49.4	9.5	16.2	24.8	3.16	1.28
18	Transitions and Attendance Verification	0.0	27.6	4.7	22.6	45.0	3.85	1.26
19	Positive Atmosphere	0.9	0.2	14.0	25.0	59.9	4.43	0.80
20	Supervision of Children	1.5	5.8	7.3	21.1	64.2	4.41	0.96
21	Foster Children's Independence	0.0	1.5	11.6	25.9	61.0	4.46	0.76
22	Supporting the Development of Self-Esteem	0.2	8.6	25.2	40.3	25.6	3.83	0.92
23	Behavior Guidance	0.2	1.1	13.8	18.5	66.4	4.50	0.79
24	Supporting Communication and Extending Children's Learning	0.0	1.1	17.0	60.1	21.8	4.03	0.66

AQI scores observed between centers. Correspondingly, only about 32% of variance occurred between the classrooms within centers.

The finding that classrooms within centers tend to be very similar supports aggregating scores across classrooms to the program level. However, it is possible for program averages to meet expectations while still having a classroom that did not meet expectations.

The results from this phase of the study provide evidence for the structural validity of both the Infant and Toddler versions of the AQI and support the computation of a single score reflecting the quality of child care. In addition, a separate, reliable indicator of the quality of teacher–child interactions can be computed if needed.

Phase 4: Establishing Concurrent Validity

Objective

The objective of the final phase of this project was to field-test the Infant and Toddler versions of the AQI as well as to evaluate the criterion validity of the instrument by investigating its relationship to other quality measures.

Sample

The final version of the AQI instrument was administered in 48 infant and 47 toddler classrooms located in 42 centers across the City of Toronto. Rooms were selected by city staff based on their having multiple infant and/or toddler rooms to maximize efficiency of data collection. Project RAs

followed the QAAs on their scheduled administrations of the Infant and Toddler versions of the AQI as part of their operational data collection. The RAs administered the instruments used to establish convergent and discriminant validity. As mentioned earlier, programs were randomly assigned to QAAs each year to minimize bias. The use of this relatively small sample in this phase was determined by time constraints for operational implementation of the AQI instrument.

Instruments

A number of instruments were used to establish convergent and discriminant validity of the AQI Infant and Toddler instruments. These measures were selected because they are frequently used to measure ECEC quality. Measures of both structural and process quality were included. These are described briefly here.

ITERS-R

The ITERS-R (Harms, Cryer, & Clifford, 2003) was developed to assess the classroom environment in child care centers. The ITERS-R is a 39-item inventory that provides a global measure of infant- or toddler-age (i.e., birth to 30 months of age) classroom environments. The items are categorized according to seven subscales: (a) Space and Furnishings, (b) Personal Care Routines, (c) Listening and Talking, (d) Activities, (e) Interactions, (f) Program Structure, and (g) Parents and Staff. Each item is scored on a 7-point scale with the following categories for each indicator: *inadequate* (1), *minimal* (3), *good* (5), and *excellent* (7). Subscale scores are then averaged to yield a mean score for each. All items can also be averaged to determine a mean classroom score. Because of its comprehensive nature, the ITERS-R is most like the AQI.

Staff–Child Ratios: Observed Data

Time stamp ratio data captured the proportion of adults to children in the participating classrooms. Observers set timers for a 30-min interval and used a standardized data collection form to record the number of children, staff, and volunteers/interns who were working in the classroom when the timers went off. This was done from the time the classroom opened and continued to be done at regular intervals throughout the morning. Because programs began and ended at different times, the number of ratio counts varied across programs. Observers collected an average of 5.84 ratio counts per classroom (range = 1–8 counts). Interns and volunteers were counted separately from staff members. Parents and visitors who were not scheduled to volunteer on the day of the observation were not included in the ratio counts.

Data Analysis Plan

All analyses for the validation phase of this study were conducted separately for the infant and toddler classroom samples to investigate the criterion validity of the measure in both types of classrooms.

Pearson (for normally distributed continuous scores) and Spearman (for skewed continuous and ordinal scores) correlations were used to investigate the convergent and discriminant validity of the Infant/Toddler AQI with other measures of structural and process quality.

Results of the correlation analyses are presented in Table 6. As can be seen in this table, both versions of the instrument showed significant, moderately strong correlations with ITERS-R total scores. Child care quality as measured by the AQI was not correlated with teacher–child ratios in the classroom. Thus, overall results of the criterion validity analysis are encouraging, especially for the Toddler AQI.

Table 6. Correlations Between Infant and Toddler AQI Scores and Other Quality Measures.

Measure	N	M (SD)	Correlation With Infant AQI	N	M (SD)	Correlation With Toddler AQI
AQI total score	48	4.0 (.41)		47	4.0 (.46)	
Structural quality						
Average child–staff ratio	46	2.2 (.50)	−.063	45	3.5 (.93)	.086
Process quality						
ITERS-R total score	48	4.1 (.59)	.322*	45	4.2 (.62)	.402**

Note. Ratio scores represent the number of staff to children and do not include volunteers, as volunteer involvement in classrooms is transient. AQI = Assessment for Quality Improvement; ITERS-R = Infant/Toddler Environment Rating Scale–Revised.
*$p < .05$. **$p < .01$.

Discussion

An increasing number of children are attending child care centers. More important, the quality of the care children experience in these centers has a significant bearing on their developmental trajectories. As a result, the ability to accurately and efficiently measure global classroom quality in child care centers for research and applied purposes has been, and continues to be, of crucial importance. This study involved a rigorous and labor-intensive measure development and testing process for the Infant and Toddler center classroom versions of the AQI. The AQI is used as part of the City of Toronto's QRIS. Given the high-stakes context of this accountability system for ECEC programs it is important that the measures be well developed and efficient not only to administer in a cost-effective manner but also to capture what is most critical in monitoring the care provided to children in their early years.

Summary of Results Across all Four Phases of the Study

In the first phase of the study, input about the Infant and Toddler AQI measures was gathered from more than 300 focus group participants who represented different stakeholders in the ECEC sector. This led to many modifications that were trialed through an iterative process of repeated field-testing. In the second phase, IRT modeling was conducted on more than 240 randomly selected infant and toddler classrooms in Toronto. This led to numerous modifications that ensured that the ordering of subitems was logical and coherent. This provided us with empirical guidance to make modifications that ensured that items at the lower end of the scale were empirically shown to be easier to attain than items at the high range of the scale. IRT modeling is an important early step in the measure development process that is often overlooked. For example, we are only aware of a very recent study on the Early Childhood Environment Rating Scale–Revised using IRT modeling (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2013) that found many issues with the ordering of this well used measure. In the third phase of this study, a large sample of Infant and Toddler AQI assessments were analyzed to determine the psychometric properties of the AQI. Here too results were encouraging in that both measures hung together well. Several other important findings from this phase of the study were that the AQI total score can be used to represent global classroom quality and that the Interactions section of the measure can be used as a stand-alone measure. In addition, we found empirical support for the city's reporting strategy, which involves aggregating AQI scores across classrooms that serve the same age groups. Finally, the fourth phase of the study provided a first round of support for the concurrent validity of the AQI versions we explored. Specifically, we found correlations of moderate magnitude between both the Infant and Toddler versions of the AQI and the ITERS-R. We did not find correlations between the Infant and Toddler versions of the AQI and structural aspects of quality. Research on associations between structural and process quality has been inconsistent (Dowsett, Huston, Imes, & Gennetian, 2008; Phillips, Voran, Kisker, Hower, & Whitebook, 1994), so our findings are in keeping with many in the literature. Overall, given

that the administration of the AQI takes a fraction of the time required to administer the ITERS-R, this level of correspondence is encouraging.

What Did We Learn About the Quality of ECEC Programs in Toronto?

Currently, very few data exist on infant and toddler classroom quality in Canada, and none that we know of exist for Toronto. Although not the primary focus of this article, Phase 4 provided us with quality data based on approximately 100 classrooms in centers that participated in the City of Toronto's QRIS.

Structural Quality

The average child–staff ratios observed in Phase 4 of the study were 2.2:1 and 3.5:1 for infant and toddler classrooms, respectively. According to the Day Nurseries Act (Ontario Day Nurseries Act, R. S.O. 1990, c. D-2. Revised Regulations of Ontario, Regulation 262), required child–staff ratios are 3.33:1 for infant rooms and 5:1 for toddler rooms. Thus, the average ratios observed in this sample exceeded the minimum required ratios.

Process Quality

The average AQI score from Phase 4 was 4.0 in both infant and toddler classrooms. Because the city considers any score below 3 to be a failure to meet expectations, it is not surprising that the average score fell between 3 and 5. Average ITERS-R scores were 4.1 in infant rooms and 4.2 in toddler rooms. The ITERS-R is scored on a 7-point scale, and a score of 4 falls between what is considered *minimal* and *good quality*. It is noteworthy that these scores are comparable to (Setodji, Le, & Schaack, 2013) or higher than (Scarr, Eisenberg, & Deater-Deckard, 1994) those reported in the literature. Thus, there appears to be substantial room for improvement in quality in ECEC programs in Toronto, and findings highlight areas that could be targeted for improvement through staff development opportunities. Clearly there is a need for continued support for QRIS and other policy levers aimed at ongoing monitoring and quality improvement efforts.

Limitations

One of the major limitations of our study is the limited data available to test the concurrent validity of the AQI. Future studies should include larger samples and a broader range of measures, including the CLASS, with its in-depth coding of the quality of staff–child interactions. However, it is important to note that unlike the CLASS, the AQI is intended as a global measure of classroom quality that captures interactions, aspects of the physical environment, health and safety factors, and so on. Thus, the most appropriate comparison measure for the AQI is the ITERS-R. Future research should also include measures of child outcomes to establish the predictive validity of the AQI measures.

Implications of Our Findings

Across all phases of the study, the psychometric properties of the Infant and Toddler versions of the AQI as well as the face and construct validity appeared strong. As a result of the rigorous process adopted in this study we have greater confidence that the content of the AQI captures what it sets out to capture and that it does so in a reliable and efficient way. Thus, results from this first study of the Infant and Toddler versions of the AQI are very encouraging. Perhaps the biggest contribution made by the AQI is its relative efficiency compared to existing measures of global classroom ECEC quality. Two classrooms can be scored in 2–3 hr using the AQI (i.e., a minimum of 1 hr per room). When the ITERS-R is used, scoring two classrooms requires 6–10

hr. This enormous difference is especially meaningful given efforts to use ERS for applied purposes as part of QRIS. At the moment, different localities make different compromises. For example, in North Carolina ERS are administered only once every 3 years as part of the state's QRIS (National Center on Child Care Quality Improvement, 2012). Although this is understandable given limited resources, reducing the frequency of measurement to cope with constrained resources is especially worrisome when the goal of the assessments is to inform and maintain ongoing quality improvements. Thus, in conclusion, the Infant and Toddler versions of the AQI hold promise as efficient, empirically based measures of quality in ECEC classrooms that serve the youngest children.

In addition to acting as an assessment tool, the AQI has served as a resource tool, educating ECEC staff on aspects of structural and process quality that are considered to be important in ECEC settings. This has influenced short- and long-term program planning and goal setting and assisted educators in creating environments that are conducive to children's cognitive, academic, and social-emotional development. Furthermore, the AQI is a helpful resource for families, helping them make informed decisions about child care for their children.

Acknowledgments

We would like to thank the many center directors, supervisors, and staff who made this study possible. We also want to thank the numerous city staff, starting with Elaine Baxter-Trahair, who facilitated every aspect of this study. In addition, we would like to thank Anastasia Howe-Bukowski and Sharon Pauker and the amazing research team that helped collect the data presented in this article.

References

Burchinal, M. R., Roberts, J. E., Nabors, L. A., & Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development*, *67*(2), 606–620. doi:10.2307/1131835

Bushnik, T. (2006). *Child care in Canada: Child and youth research paper series* (Catalogue No. 89-599-XIE). Retrieved from the Statistics Canada website: http://publications.gc.ca/Collection/Statcan/89-599-MIE/89-599-MIE2006003.pdf

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*(3), 579–620.

Carpenter, M., Martin, M., & Russell, S. (2005). *Who's caring for our babies now? Revisiting the 2005 profile of early care and education for children birth to three in North Carolina*. Retrieved from the Child Care Services Association website: http://www.childcareservices.org/_downloads/research/IT_State%20report_08.pdf

Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly*, *17*(1), 52–86. doi:10.1016/S0885-2006(02)00133-3

Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly*, *23*(1), 69–93. doi:10.1016/j.ecresq.2007.06.003

Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*(5), 1454–1475. doi:10.1111/1467-8624.00617

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.

Fisher, M. A. (2011). The early years and brain development: The critical connections of building relationships. In G. L. Ensher & D. A. Clark, *Relationship-centered practices in early childhood: Working with families, infants, and young children at risk*, (pp. 33-44). Baltimore, MD: Brookes Publishing.

Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., & LaGrange, A. (2006). Towards a predictive model of quality in Canadian child care centers. *Early Childhood Research Quarterly*, *21*(3), 280–295. doi:10.1016/j.ecresq.2006.07.005

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, *49*(1), 146–160. doi:10.1037/a0027899

Graves, S. L., Jr., & Howes, C. (2011). Ethnic differences in social-emotional development in preschool: The impact of teacher child relationships and classroom quality. *School Psychology Quarterly*, *26*(3), 202–214. doi:10.1037/a0024117

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625–638. doi:10.1111/1467-8624.00301

Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale–Revised*. New York, NY: Teachers College Press.

Hestenes, L. L., Kintner-Duffy, V., Wang, Y. C., La Paro, K., Mims, S. U., Crosby, D., … Cassidy, D. J. (2015). Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children. *Early Childhood Research Quarterly*, *30*, 199–214. doi:10.1016/j.ecresq.2014.06.003

Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology*, *24*(1), 53–57. doi:10.1037/0012-1649.24.1.53

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, *26*(3), 395–400.

Jeon, L., & Buettner, C. K. (2015). Quality rating and improvement systems and children's cognitive development. *Child & Youth Care Forum*, *44*(2), 191–207. doi:10.1007/s10566-014-9277-7

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage.

Landry, S. H., & Smith, K. E. (2010). Early social and cognitive precursors and parental support for self-regulation and executive function: Relations from early childhood into adolescence. In B. Sokol, U. Müller, J. Carpendale, A. Young, & G. Iarocci (Eds.), *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 386-417). New York, NY: Oxford University Press.

McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology*, *20*(2), 244–260. doi:10.1037/0012-1649.20.2.244

Mitchell, A. (2005). *Stairs steps to quality: A guide for states and communities developing quality rating systems for early care and education*. Retrieved from the Early Childhood Finance website: http://www.earlychildhoodfinance.org/downloads/2005/MitchStairSteps_2005.pdf

National Center on Child Care Quality Improvement. (2012). *QRIS in statutes and regulations* (No. 457). Retrieved from http://qrisnetwork.org/sites/all/files/resources/gscobb/2012-04-02%2011%3A56/Report.pdf

National Early Childhood Accountability Task Force. (2007). *Taking stock: Assessing and improving early childhood learning and program quality*. Retrieved from http://policyforchildren.org/wp-content/uploads/2013/07/Taking-Stock.pdf

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, *71*(4), 960–980. doi:10.1111/cdev.2000.71.issue-4

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, *13*(3), 199–206. doi:10.1111/1467-9280.00438

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology*, *39*(3), 451–469. doi:10.1037/0012-1649.39.3.451

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly*, *43*(3), 451–478.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., & Kagan, S. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*(5), 1534–1553. doi:10.1111/1467-8624.00364

Phillips, D., Voran, M., Kisker, E., Hower, C., & Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? *Child Development*, *65*(2), 472–492. doi:10.2307/1131397

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual: K-3*. Baltimore, MD: Brookes.

Read, V. (2014). *Developing attachment in early years settings: Nurturing secure relationships from birth to five years*. New York, NY: Routledge.

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, *9*(2), 131–151.

Schore, A. (1994). *Affect regulation and the origin of the self: The neurobiology of emotional development*. Mahwah, NJ: Erlbaum.

Setodji, C., Le, V., & Schaack, D. (2013). Using generalized additive modeling to empirically identify thresholds within the ITERS in relation to toddlers' cognitive development. *Developmental Psychology*, *49*(4), 632–645. doi:10.1037/a0028738

Squires, J. (2012). *Assessing young children's social and emotional development*. Baltimore, MD: Brookes.

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations*. Retrieved from the Office of Planning, Research, and Evaluation website: http://www.acf.hhs.gov/sites/default/files/opre/qrs_compendium_final.pdf

Whitebook, M. (2003). *Early education quality: Higher teacher qualifications for better learning environments—A review of the literature*. Retrieved from the Center for the Study of Child Care Employment website: http://www.irle.berkeley.edu/cscce/wp-content/uploads/2003/01/Early_Ed_Quality.pdf

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W., … Zaslow, M. J. (2013). *Investing in our future: The evidence base for preschool education*. Retrieved from the Foundation for Child Development website: fcdus.org/sites/default/files/Evidence Base on Preschool Education FINAL.pdf

Zellman, G., Perlman, M., Le, V., & Setodji, C. (2008). *Assessing the validity of the Qualistar early learning quality rating and improvement system as a tool for improving child-care quality*. Retrieved from the Rand Education website: http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG650.pdf

Appendix

Scoring of the Assessment for Quality Improvement (AQI)

The AQI is a global measure of classroom quality. More specifically, the AQI looks at the daily program plan staff use to plan and document programming for the week; displays that are available within the classroom; aspects of the indoor learning environment; materials and learning opportunities that are available in the areas of art, science, literacy, and math; as well as opportunities for students to develop social and emotional skills. The AQI also includes items that assess health and safety, including diapering and hand-washing routines, cribs and bedding, and transition and attendance protocols.

Classrooms are rated along a continuum, with a score of 1 or 2 defined as *inadequate* or *not meeting expectations*, 3 defined as *meeting expectations*, and 4 or 5 as *exceeding expectations*. Scoring begins in the "Does not meet expectations" column. A check in this column is considered noncompliance. Thus, to move up to the "Meets expectations" category, a classroom must not exhibit any of the characteristics listed under the "Does not meet expectations" category. Similar to Environment Rating Scales ratings, a classroom must meet all of the subitems within a given scoring level before it is able to proceed to the next level of scoring. If classrooms only meet 50% of the subitems within a given category, they are unable to move forward and are assigned that score. For example, to have a score of 4 in the Activities and Experiences domain, a classroom should not have any checks in the column "Does not meet expectations," have all of the checks in the column "Meets expectations," and have at least 50% of checks in the column "Exceeds expectations" (see Figure A-1). A room's domain score is determined by its performance across each of the levels on each domain. A room's AQI total score is calculated as an average across all domains. The AQI is a much more efficient tool than other quality assessment tools, taking a minimum of 2–3 hr to assess two rooms.

	Does not meet Expectations (1)or (2)	Meets Expectations (3)	Exceeds Expectations (4)or (5)	
3. Activities and Experiences	❑ Activities and experiences offered do not reflect the current program plan ❑Activities and experiences offered are not developmentally appropriate ❑ Activities do not promote choice for children.	❑ There is current documentation which demonstrates that observations of children are used in the development of activities ❑ Activities are adapted to meet any child's individual needs [2] ❑ Standardized Developmental Screening tool completed for all children [2] ❑Evidence of opportunities to discuss developmental progress with families. [2]	❑Picture documentation of activities available [2] ❑Activity resources accessible for families ❑Portfolios regarding each child's development are accessible to families ❑Enrichment program in addition to regular program is included monthly ❑Staff provide spontaneous resources to allow the child to follow their own learning path.	1 2 3 4 5

Figure A-1. Structure of the Activities and Experiences domain.