Child Care Health Consultation Improves Infant and Toddler Care

Rosemary Johnston, RN, BSN, MSN, Beth A. DelConte, MD, FAAP, Libby Ungvary, MEd, Richard Fiene, PhD, & Susan S. Aronson, MD, FAAP

ABSTRACT

Introduction: Many families enroll their infants and toddlers in early education and child care programs. The Pennsylvania Chapter of the American Academy of Pediatrics recruited 32 child care centers that care for infants and toddlers to be linked with a child care health consultant (CCHC).

Method: Project staff assigned the centers alternately to an immediate intervention or a 1-year delayed intervention (contrast) group. At entry into the project, and then 1 and 2 years later, an evaluator assessed center compliance with 13 standards for infants and toddler care selected from Caring for Our Children: National Health and Safety Performance Standards (3rd ed.). Project staff linked the Immediate Intervention centers with a CCHC in Year 1. In Year 2, in a crossover comparison, project staff linked Contrast centers with a CCHC.


KEY WORDS

Child care, child care health consultation, health and safety, infants and toddlers

INTRODUCTION

Nationally, about 48% of children younger than 3 years of age are enrolled in organized child care facilities (Laughlin, 2013). Early educators (child care staff) care for these children for many hours and many days. The quality of their care has lifelong impact on their physical, developmental, and social-emotional well-being (Garcia, Heckman, Leaf, & Padros, 2016).

In 2013, the Early Childhood Education Linkage System (ECELS), a program of the Pennsylvania (PA) Chapter of the American Academy of Pediatrics (AAP)...
received a 3-year grant from the Maternal and Child Health Bureau (MCHB). The purpose of the grant was to “improve state infant/toddler [I/T] child care quality initiatives (Quality Rating and Improvement Systems [QRIS] and professional development)...” MCHB’s grant required selection and promotion of 10 or more standards from a list provided by MCHB from Caring for Our Children: National Health and Safety Performance Standards, Guidelines for Early Care and Education Programs, 3rd ed. (CFOC3; AAP, American Public Health Association, & National Resource Center for Health and Safety in Child Care and Early Education, 2011).

Child care programs in PA’s QRIS, called Keystone STARS, are ranked from the entry level at STAR 1 to STAR 4. To earn a rating, programs must comply with state regulations and meet the requirements listed for the designated STAR level on the PA Key Web site (www.pakeys.org). For a STAR 4 rating, a center that serves infants and toddlers must have scores at or above 5 (good) on the seven subscales of the Infant and Toddler Environment Rating Scale–Revised Edition (ITERS-R; Harms, Cryer, & Clifford, 2006). The Personal Care Routines subscale of the ITERS-R has some health and safety items. Scores in this subscale and on health and safety items in some of the other subscales are among the lowest scoring ITERS-R items in PA and elsewhere. This finding is reported by the PA Key Program Quality Assessment Team (2016) and by the authors of the ITERS-R (Harms and Cryer, personal communication, 2014).

Child care health consultants (CCHCs) use observation, education, collaborative decision making, coaching, and mentoring to achieve quality improvement in the QRIS (Zaslow, Tout, & Halle, 2012). CCHCs base their work on needs and feasible implementation. For more than a decade, published research has confirmed that child care health consultation is an effective approach to improving health and safety compliance with national child care standards (Alkon & Bernzweig, 2008; Alkon et al., 2008; Alkon, Bernzweig, Kim, Wolff, & Mackie, 2009; Alkon et al., 2014; Alkon et al., n.d.; Alkon, Sokal-Gutierrez, & Wolff, 2002; Banghart & Kraeder, 2012; Carabin et al., 1999; Crowley, 2006; Isbell et al., 2013; Moon & Oden, 2005; Organizational Research Services & Geo Education and Research, 2007; Pacific Research and Evaluation, 2007, 2008; Ramler, Nakatsuikawa-Ono, Loe, & Harris, 2006; Roberts et al., 2000a, 2000b). Most of these studies did not specifically target care for infants and toddlers.

Published studies document the following specific improvements associated with involvement of a CCHC. Sanitation and hygiene reduced respiratory and gastrointestinal illness and days absent for illness among young children in group care (Carabin et al., 1999; Kotch et al., 2007; Roberts et al., 2000a, 2000b). Nationally recommended practices related to active play, nutrition, and food handling were adopted (Alkon et al., 2014). Policies and procedures accompanied by staff training reduced hazards and injuries (Kotch, 2002; Organizational Research Services & Geo Education and Research, 2007). Training about safe infant sleep positioning and the infant sleep environment reduced risk of sudden infant death syndrome (Moon & Oden, 2005). Better monitoring and tracking of immunization data in child care programs was associated with more children having up-to-date vaccine documentation (Alkon & Bernzweig, 2008).

The PA AAP established ECELS in 1989. ECELS maintains a CCHC Registry and regularly communicates with registered CCHCs to provide professional development, technical assistance, and tools to enable their implementation of the CCHC role. PA’s CCHCs include private and public health service providers and health professionals who teach in academic settings. Funding for CCHC work is unpredictable, making recruitment, education, and retention of CCHCs challenging.

PA’s child care regulations require that child care providers have documents showing that enrolled children are up to date with preventive health services recommended by the AAP, including “a review of the child’s immunized status according to recommendations of the ACIP [Advisory Committee on Immunization Practices]” (PA Department of Human Services, 2008). This regulation is not enforced. Few providers use any reliable way to ensure that enrolled children are up to date. ECELS encourages child care centers to use a well-tested and routinely updated online software application called WellCareTrackerTM (Weinburg, 2002) to check child health records for up-to-date routine preventive health services. It is described, demonstrated, and offered for subscription at www.wellcaretracker.org. Using WellCareTrackerTM eases the burden for child care providers to comply with the regulation and remind families to obtain these services in a timely manner.

**METHODS**

**Design**

The PA AAP’s MCHB-funded Infant-Toddler Quality Improvement Project (ITQIP) was conducted by ECELS using a randomly assigned clinical trial with a crossover comparison of centers assigned to an immediate intervention or delayed intervention (comparison) group. ECELS (a) assessed child care center practices related to I/T care for 13 selected CFOC3 standards (AAP et al., 2011) and (b) assessed whether compliance with these practices improved when centers were linked with a CCHC.
Selection of the CFOC3 standards addressed in ITQIP

With input from early care and education stakeholders, ECELS chose 13 CFOC3 standards (AAP et al., 2011) from a list provided by MCHB (Box 1). The selection criteria were that the standard is (a) associated with the highest and most common risks of harm to I/T (AAP, American Public Health Association, & National Resource Center for Health and Safety in Child Care and Early Education, 2013), (b) measurable and amenable to improvement with technical assistance and professional development provided by a CCHC over a 12-month period, and (c) found by state inspectors to have a high level of noncompliance according to state data (PA Office of Child Development and Early Learning, 2010).

Evaluation plan

The evaluation plan is a classic randomly assigned crossover clinical trial. See Figure 1 for the evaluation plan logic model.

The ITQIP staff and consultants developed the evaluation tool described below. The ITQIP Project Coordinator (first author) and the evaluators collected data from participating centers at three points: when centers enrolled in the study (Pretest) and then 1 year (Posttest 1) and 2 years later (Posttest 2). One of the consultants (fourth author) compared the two groups on the pretest for equivalency and then on each of the two posttests. These data are discussed in the Results: Immediate Intervention Versus Delayed Intervention (Contrast) Group section. One year after the pretest data were collected, the participating centers were switched to a crossover comparison format. At this point, ITQIP ended the subsidy for the CCHCs who were working with the centers in the immediate intervention group and provided the subsidized CCHC linkage to the centers in the delayed intervention (contrast) group.

When a center enrolled in ITQIP, the ITQIP coordinator interviewed the center director by phone. She gathered demographic data, including the number of enrolled I/Ts, where and when I/T activities occurred in the center, and the number of children who met the MCHB definition of special health needs. She asked the director to submit up to five of any care plans the center had for these children, redacted for confidentiality. The MCHB definition of a child with special health care needs is noted in CFOC3 standard 3.5.0.1 as “a child who has or is at increased risk for chronic physical, developmental, behavioral or emotional conditions and who requires health and related services of a type or amount beyond that required by children generally” (AAP et al., 2011).

The ITQIP coordinator selected the rooms for the evaluator to observe as those with the largest number of children in the age group. The evaluators recorded observations in one infant and one toddler room at each center.

The evaluator collected a random sample of immunization records for up to 10 infants and 10 toddlers with the names redacted for confidentiality. The ITQIP coordinator used WellCareTracker™ software to check these immunization records. The ITQIP coordinator evaluated the care plans that the director submitted for the presence of the appropriate components from the list of the 14 components specified in CFOC3 standard 3.5.0.1 (AAP et al., 2011) and a 15th component, the presence of the health care provider’s signature, that is required by PA regulations (Box 2).

BOX 1. CFOC3 standards chosen for ITQIP

1.4.5.2 Child Abuse and Neglect Education
3.4.4.1 Recognizing and Reporting Suspected Child Abuse, Neglect, and Exploitation
2.1.2.1 Personal Caregiver/Teacher Relationships for Infants and Toddlers
2.2.0.2 Limiting Infant/Toddler Time in Crib, High Chair, Car Seat, and other restraining equipment
3.1.3.1 Active Opportunities for Physical Activity
3.1.4.1 Safe Sleep Practices and SIDS Risk Reduction
3.2.1.4 Diaper Changing Procedure
3.2.2.1 Situations That Require Hand Hygiene
3.2.2.2 Handwashing Procedure
3.6.3.3 Training of Caregivers/Teachers to Administer Medication
3.5.0.1 Care Plan for Children with Special Health Care Needs
5.4.5.2 Cribs
7.2.0.1 Immunization Documentation

Note. CFOC3, Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.); ITQIP, Infant-Toddler Quality Improvement Project; SIDS, sudden infant death syndrome.
The ITQIP coordinator scored the evaluator’s observations of diapering, hand hygiene, and medication administration. She promptly prepared a summary of all the findings for the center and sent the summary to the center director and the linked CCHC before the first CCHC site visit. The summary delineated areas of strengths and areas to improve based on the evaluation tool results. To facilitate use of the data by the center staff and CCHCs, the summary included the text of the evaluation tool item, the center’s score on the item, and the reason why the center met or did not meet the standard. The CCHC contacted the center within 2 weeks after receiving the summary to set up the initial site visit.

**Evaluation Tool**
The ITQIP staff prepared the items on the evaluation tool from performance guidelines specified in the 13 selected CPOC3 standards (AAP et al., 2011). ITQIP consultants (fourth and fifth authors) and the ECLES staff reviewed the tool for clarity and validity of content. After several rounds of revisions, the ITQIP coordinator and a prospective ITQIP evaluator field-tested the tool, further revised it, and then field-tested it again, this time testing for interrater reliability with two evaluators independently and simultaneously using the tool.

The ITQIP evaluation tool has four sections: (a) Demographic Information collected in the phone interview (35 items), (b) Observations (64 items), (c) Interview Questions (28 items), and (d) Documents (14 items). The score awarded to items on the evaluation tool was based on the criteria listed in Box 3. A score of 2 or 3 for an item was considered a strength, and a score of 0 or 1 for an item was considered an area to improve. This total score was the sum of the scores for each item. The total number of scorable items on the evaluation tool is 106, with a maximum score of 318. The documents assessed include training records, written policies, care plans for children with special needs, immunization data, and PA child abuse clearances.

ITQIP assigned each scorable item to one of the 10 topic areas addressed by the 13 CPOC3 standards selected for the project (AAP et al., 2011). See Table 1.

**Sampling design: Recruitment, random assignment, and retention of centers**
ECELS recruited Keystone STAR 2 and STAR 3 centers by distributing a flyer about the project. Programs with higher STARS ratings qualify for higher payments for children whose care is state subsidized. The highest payments are for children enrolled in STAR 4 centers. The increased payment for a higher rating is a quality improvement incentive. Also, ECELS offered participating centers three free $10 credit-awarding reviews for ECELS self-learning modules. The flyer was included in the newsletters of a variety of organizations: four of the five regional state-supported sources of professional development (Regional Keys), the PA Child Care Association, the Pittsburgh Association for the Education of Young Children, and United Way. Because the northwestern region of the state has the fewest centers, recruitment from that region was not attempted.

As the centers joined ITQIP, the project coordinator assigned them alternately to one of two groups, either the immediate intervention group or the delayed intervention (contrast) group. ITQIP enrolled centers from all four targeted regions of the state.

**BOX 2. Care plan components evaluated for children with special needs**

| 1. A list of the child’s diagnoses |
| 2. Contact information for the child’s health care provider and any subspecialists |
| 3. Medications to be administered on a scheduled basis |
| 4. Medications to be administered in an emergency with clearly stated signs and symptoms in lay language about when to give the medication |
| 5. Procedures to be performed while in care |
| 6. Allergies |
| 7. Diet modification that the child requires |
| 8. Activity modifications |
| 9. Environmental modifications |
| 10. Triggers that cause a reaction to avoid |
| 11. Symptoms for caregivers/teachers to observe |
| 12. Behavioral modifications beyond those needed for a typically developing child |
| 13. Emergency response plans for a facility emergency and if the child has an emergency event |
| 14. Special skills training and education required and provided for the staff |
| 15. Health care provider signature (required by Pennsylvania regulation) |

Note. **Fourteen components specified in the Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.) standard 3.5.0.1.** (American Academy of Pediatrics et al., 2011) and a 15th required by Pennsylvania child care regulation.
Centers enrolled in ITQIP agreed to

- allow a 4- to 5-hour site evaluation once a year for 3 years,
- work with a CCHC for a period of 1 year to improve I/T health and safety,
- accept random assignment to one of the two project groups,
- provide access to redacted immunization records and care plans for evaluation,
- pay $240.00 of the $500 honorarium ITQIP paid to their CCHC, and
- remain in ITQIP for 3 years.

Recruitment and roles of evaluators and CCHCs

ITQIP recruited 17 evaluators from the list of CCHCs who had previously received CCHC training from ECELS and from the nurses in the Maternal Infant and Early Childhood Home Visiting Program. All evaluators were health professionals with pediatric experience related to observed items. Most had experience working with CFOC3 standards (AAP et al., 2011). The evaluators learned how to use the evaluation tool by participating in a live Webinar or by using the recording of the Webinar. All evaluators received a copy of the evaluation tool and a training manual with instructions for completing the evaluation. Seven evaluators were also CCHCs in this project. None of the evaluators who were CCHCs in ITQIP were linked with centers they evaluated.

The evaluators gave their completed evaluation tools to the ITQIP coordinator to score and summarize. The coordinator reviewed each submitted evaluation tool and then discussed the documentation with the evaluator by phone to make sure the scoring was as intended.

Child Care Health Consultants.

ECELS recruited 14 registered nurses and one physician as CCHCs. The ITQIP coordinator (first author) has worked as a CCHC for more than 15 years. She and the project’s director and primary investigator, a pediatrician (second author) educated, coached, mentored, and supported the work of the CCHCs. The CCHCs participated in a Webinar about the project scope and the use of the selected CFOC3 standards (AAP et al., 2011). They received a training manual that included the 13 selected CFOC3 standards (AAP et al., 2011) and resources to support best practice in each of the 10 topic areas. ITQIP provided additional resources and periodic CFOC3 updates (AAP et al., 2011).

During the site visit, the CCHC compared her observations with those in the summary and solicited concerns about health and safety practices from the center’s staff. Then the director, program staff, and CCHC chose three of the 10 topics as the primary focus of the center’s improvement. The CCHC helped the center staff prepare an action plan to work on the three topic areas they chose. Action plans included filling gaps in knowledge, developing policies for staff and family handbooks, and improving staff practices. The CCHCs and center directors arranged all subsequent contacts and visits over the next 12 months.

Quarterly, the CCHCs sent the ITQIP coordinator documentation of their work and progress toward goals. The CCHCs submitted the center’s initial action plan and a final action plan at the end of the year that showed what the center accomplished. ITQIP paid $250 to the CCHCs upon receipt of the center’s initial action plan and date of the first CCHC visit. ITQIP paid the CCHCs an additional $250 after they submitted the final action plan from their 12-month linkage. Throughout the project, the ITQIP coordinator reviewed quarterly encounter forms that the CCHCs submitted to describe their work with the centers. This enabled the ITQIP coordinator to suggest ways to promote progress on action plans, including use of relevant health and safety resources.

**RESULTS**

**Descriptive Report**

ITQIP linked CCHCs with 32 centers. Of these, 16 centers were in the immediate CCHC-linked group, and 16 were in the delayed CCHC-linked group. In all,

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Topic areas</th>
<th>Number of items to score per topic*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>Preventing Child Abuse</td>
<td>13</td>
</tr>
<tr>
<td>PR</td>
<td>Personal Relationships</td>
<td>9</td>
</tr>
<tr>
<td>LA</td>
<td>Limited Physical Activity of Infants</td>
<td>3</td>
</tr>
<tr>
<td>AO</td>
<td>Active Opportunity for Physical Activity</td>
<td>22</td>
</tr>
<tr>
<td>SS</td>
<td>Safe Sleep Practices/SIDS Prevention</td>
<td>19</td>
</tr>
<tr>
<td>MA</td>
<td>Medication Administration</td>
<td>8</td>
</tr>
<tr>
<td>DC</td>
<td>Diaper Changing Procedure</td>
<td>16</td>
</tr>
<tr>
<td>HH</td>
<td>Hand Hygiene</td>
<td>8</td>
</tr>
<tr>
<td>IM</td>
<td>Immunization</td>
<td>3</td>
</tr>
<tr>
<td>SN</td>
<td>Care Plans for Children With Special Needs</td>
<td>5</td>
</tr>
</tbody>
</table>

Note. SIDS, sudden infant death syndrome.

*See the narrative for an item-by-item explanation of those items with significance levels (p values) based on the t tests performed on each item.
59 directors, 3481/T teachers and 1,490 infants and toddlers were directly involved in ITQIP. Three centers from each group dropped out, leaving 13 centers in each group at the completion of the project (Table 2). Over the 1-year period of CCHC linkage, 12 of the 32 programs had turnovers of two to four directors. This change in center leadership made the CCHCs’ work to improve I/T care very difficult. For the immediate intervention group, three of the original 16 centers withdrew from the project. One center in the delayed intervention (contrast) group closed during the project period; two others withdrew from ITQIP. Some centers dropped out because they were so overwhelmed with maintaining ratios in classrooms and staffing issues that they believed they could not focus on their action plans.

This report compares pretest, Posttest 1 and Posttest 2 scores for the 13 immediate intervention sites and 13 delayed intervention (contrast) sites that remained enrolled in ITQIP for the full 3 years.

ITQIP did not require a specific time spent in the CCHC role for each linkage. The CCHCs in the immediate intervention group provided an average of 14 hours of consultation per site (range = 2.25–28.75 hours). The CCHCs in the delayed intervention (contrast) group provided an average of 12.5 hours of consultation per site (range = 2–32 hours). The CCHCs completed quarterly encounter forms to report the total hours of services to their linked center, including a checklist of onsite, phone, and e-mail services. The most common CCHC interactions with centers included providing health education for the director and staff, onsite consultation at the facility, technical assistance by phone or e-mail, providing print or audiovisual materials, helping the facility comply with state regulations, and developing health policies and procedures.

Topics chosen by the centers in the immediate intervention group and the delayed intervention (contrast) group and the number of centers that chose each topic are shown in Table 3.

### Quantitative Comparison of Evaluation Tool Scores on the Pretest Versus the Two Posttests

The scores used in the quantitative comparisons are the sum of all scores on the Evaluation Tool, not only those for the topics that the center chose for special focus (Table 4).

#### Immediate intervention group

On the pretest, the range in scores was 175 to 267, with an average score of 212 out of a possible 318 points (66%). On Posttest 1, the range in scores was 213 to 297, with an average score of 254 out of a possible 318 points (79%). This change from the pretest to Posttest 1 was statistically significant ($t = −4.62, p < .0001$). Posttest2 did not show any significant change from the average score on Posttest 1, showing that the initial results from the intervention were sustained in the next year (254 to 254).

#### Delayed intervention (contrast) group

On the pretest, the range in scores was 164 to 271, with an average score of 218 out of a possible 318 points (68%). On Posttest1, the range in scores was 149 to 257, with an average score of 221 out of a possible 318 points (69%). These changes from the pretest to Posttest 1 were not significant. Posttest2 showed significant change in the average score from Posttest 1 (221 points) to Posttest 2 (243 points; $t = −1.80, p < .08$) a year after this delayed intervention (contrast) group had received their CCHC linkage.

#### Immediate Intervention Versus Delayed Intervention (Contrast) Groups

The comparison of the average scores between the Immediate Intervention (212) and Delayed Intervention (Contrast, 218) groups on the pretest was not significant, showing that the groups were equivalent. The difference between the average scores of the immediate intervention (254) and delayed intervention (contrast, 221) groups on Posttest1 was statistically significant ($t = −3.46, p < .002$), showing the effectiveness of the CCHC intervention for the immediate intervention group. Posttest2 showed no significant difference between the change in the average postintervention scores for the immediate intervention group 12 months after their CCHC-subsidized linkage and the delayed intervention (contrast) group (254 vs. 243) at the end of their 12 months of CCHC-subsidized linkage. See Figure 2 for the crossover comparison results.

---

**TABLE 2. Location and retention of recruited centers**

<table>
<thead>
<tr>
<th>Region of Pennsylvania</th>
<th>Immediate intervention group</th>
<th></th>
<th></th>
<th>Delayed intervention group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Centers recruited</td>
<td>Centers dropped out</td>
<td>Centers completed</td>
<td>Centers recruited</td>
<td>Centers dropped out</td>
<td>Centers completed</td>
</tr>
<tr>
<td>Southwest Region (Pittsburgh metropolitan area)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>South Central Region (Harrisburg metropolitan area)</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Northeast Region (Allentown/Bethlehem/Scranton)</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Southeast Region (Philadelphia metropolitan area)</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>3</td>
<td>13</td>
<td>16</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>
The crossover comparison results (Figure 2) show the relationship between the immediate intervention and the delayed intervention (contrast) groups in a crossover design. It clearly shows how effective the intervention (pretest to Posttest 1) was for the immediate intervention group and that the effects persisted after 1 year without a subsidized CCHC linkage (Posttest 1 to Posttest 2). It also shows that the intervention was effective when the delayed intervention (contrast) group was switched to receive the CCHC intervention with targeted training, technical assistance, and collaborative consultation a year after their pretest assessment (Posttest 1 to Posttest 2).

For the Immediate Intervention Group After 1 Year of Linkage With a CCHC
Among the items in each topic area (Table 1), the following items showed statistically significant improvement (pretest to Posttest 1).

**Medication administration**
The director had documentation that the staff who are authorized to give medications have received medication administration training within the year from a health professional ($p < .001$).

**Safe sleep**
The number of written safe sleep policies containing the required elements increased ($p < .05$). Teachers ($p < .01$) and parents ($p < .05$) reviewed the safe sleep policies and were educated about safe sleep practices ($p < .05$).

**Child abuse**
Child abuse policies contained the required elements ($p < .05$). Both infant and toddler teachers were educated about child abuse and how, as mandated reporters, they are required to personally report incidents they suspect might involve child maltreatment ($p < .001$). The number of centers having required clearance documents on file for teachers increased ($p < .05$).

**Active opportunities for physical activity**
Infants (birth through 12 months of age) were taken outside two to three times per day, as tolerated ($p < .05$). Toddlers (12 months through 3 years)

### TABLE 3. CFOC3 topics chosen by centers by intervention group

<table>
<thead>
<tr>
<th>CFOC3 topics</th>
<th>Immediate intervention group (chose each topic)</th>
<th>Contrast intervention group (chose each topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe Sleep Practice</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Medication Administration</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Child Abuse Prevention</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Care Plans for Children with Special Needs</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Diaper Changing Procedure</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Limited Physical Activity of Infants</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Hand Hygiene</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Immunization</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Personal Relationships</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Active Opportunity for Physical Activity</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. CFOC3, Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.).

### TABLE 4. Quantitative results of the evaluation from the pretest to two posttests

<table>
<thead>
<tr>
<th>Intervention group</th>
<th>Range</th>
<th>Average</th>
<th>%</th>
<th>Possible total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>175–267</td>
<td>212</td>
<td>66</td>
<td>318</td>
</tr>
<tr>
<td>Posttest 1</td>
<td>213–297</td>
<td>254</td>
<td>79</td>
<td>318</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>137–286</td>
<td>254</td>
<td>79</td>
<td>318</td>
</tr>
<tr>
<td>Delayed intervention (contrast group)</td>
<td>Range</td>
<td>Average</td>
<td>%</td>
<td>Possible total</td>
</tr>
<tr>
<td></td>
<td>164–271</td>
<td>218</td>
<td>68</td>
<td>318</td>
</tr>
<tr>
<td></td>
<td>149–257</td>
<td>221</td>
<td>69</td>
<td>318</td>
</tr>
<tr>
<td></td>
<td>170–283</td>
<td>243</td>
<td>76</td>
<td>318</td>
</tr>
</tbody>
</table>

Note. CCHC, child care health consultant.

---

$^a$Statistically significant change ($t = -4.62, p = .0001$) from pretest to Posttest 1 for the immediate intervention group after the intervention of a 1-year linkage with a CCHC.

$^b$Statistically significant change ($t = -1.80, p = .08$) from Posttest 1 to Posttest 2 for the delayed intervention group after the intervention of 1 year of CCHC linkage.

$^c$Statistically significant change ($t = -3.46; p < .002$) for Posttest 1 between the immediate intervention group and the delayed intervention (contrast) group.
went outside except in weather that poses a significant health risk \( (p < .05) \).

### Diaper changing

Before the beginning of the diaper change, changing table paper was placed over the diapering surface, followed by the gathering of supplies needed for the diaper change from the containers in which they are stored and use of gloves \( (p < .05) \).

### Hand hygiene

Observed times when toddlers \( (p < .01) \) and the toddler teachers/caregivers \( (p < .05) \) should have washed their hands showed statistically significant improvement after CCHC linkage.

### For the Delayed Intervention (Contrast) Group After 1 Year of Linkage With a CCHC

Among the items in each topic area (Table 1), the following items showed statistically significant improvement (Posttest 1 to Posttest 2).

### Safe sleep

Safe sleep policies that contained all the elements that should be in a safe sleep policy per \( CFOC3 \) standard 3.1.4.1. \( (p < .05; AAP \ et \ al., \ 2011) \). The facility had documentation that parents reviewed the center's safe sleep policy and were educated about safe sleep practices \( (p < .05) \). There was no soft or loose bedding or other objects in a crib when an infant was in the crib \( (p < .05) \). Caregivers and teachers checked on sleeping infants often enough (about every 5 minutes) to be sure that the infant was still breathing \( (p < .05) \).

### Medication administration

The name of a child to receive medication was verified before the medication was administered to that child \( (p < .05) \).

### Diaper changing

Bottom clothing was removed, including shoes and socks, if feet were unlikely to be kept from contacting soiled skin or surfaces. If clothing was soiled, it was removed and placed in a plastic bag \( (p < .05) \).

### Special needs

The number of care plans submitted that included the required elements in a care plan for children with special needs per the \( CFOC3 \) standard 3.5.0.1 increased \( (p < .05; AAP \ et \ al., \ 2011) \).

### Additional Findings of Interest

### Immunization documentation

Only one center chose to work on documentation of up-to-date immunization status as an action plan focus. Overall, the immunization data for the two groups showed low compliance with \( CFOC3 \) standard 7.2.0.1 \( (AAP \ et \ al., \ 2011) \) and PA's immunization regulations \( (PA \ Department \ of \ Human \ Services, \ 2008) \). On the pretest, in the immediate intervention centers, 22% of the immunization records for infants and 43% of the immunization records for toddlers were up to date.
The data for the two groups showed low compliance with care plans for children with special needs cords for toddlers showing up-to-date vaccines. To 38% for infants but dropped from 40% to 27% of the relayed intervention (contrast) centers improved from 25% records for toddlers were up to date. On Posttest 1 the de-

zation records for infants and 40% of the immunizations for infants, 43% for toddlers. On the pretest for the de-

Little change occurred for this group on Posttest 1 (36% to infectious diseases and injuries because of their age-appropriate behavior and abilities, their immature immune systems, and their lack of understanding of risk. Maintaining safe and healthful environments and practices involves removal of hazards and provision of policies and procedures, as well as compliance with quality standards by everyone in the group.

Numerous studies have shown the effectiveness of child care health consultation. This study focused on I/T care. The immediate intervention group showed significant improvement in policy development for safe sleep and child abuse and in education about safe sleep practices, preventing child abuse, and medication administration training. Some improvement in diaper changing and hand hygiene procedures occurred. The delayed intervention (contract) group showed significant improvement in safe sleep procedures, policies and education, medication administration procedure, diaper changing procedures, and care plans for children with special needs with appropriate information and signed by a health care provider.

The data collected by ITQIP show that many children with special needs lacked appropriate care plans. After finding little improvement in the immediate intervention group for centers having care plans with needed elements, ITQIP chose this topic as the focus of an MCHB-required continuous quality improvement initiative. ITQIP provided an audioconference for the CCHCs and gave them resources for teaching what should be in a care plan. CCHCs reported that they were most successful at helping the centers have complete, useful care plans for children with disease-specific conditions.

The areas chosen to target varied from center to center. Immunization was chosen by only one center. At the time of the study, neither regulation inspectors nor quality rating assessors were checking whether the center had documentation that the enrolled children were up to date with their vaccines. With little incentive or sanctions, documentation of up-to-date immunization status was poor.

Improvements occurred in some practices specified in selected CFOC3 standards. Many of the directors said they appreciated the help they received from the CCHCs that ITQIP linked with their centers. The director of one center, part of a corporation with centers in 12 states, advocated for improving sleep policies for all the centers in her company. This advocacy could lead to widespread improvement.

The centers that participated in this project were STAR 2 and STAR 3 programs that responded to an invitation to participate in ITQIP to improve. They were willing to contribute a modest copayment to work with a CCHC and wanted to raise their STAR rating and consequent higher payments for subsidized enrollees.

DISCUSSION AND CONCLUSIONS
Quality early education and child care have been shown to be associated with lifelong benefits (Garcia et al., 2016). Young children are especially vulnerable to infectious diseases and injuries because of their age-appropriate behavior and abilities, their immature immune systems, and their lack of understanding of risk. Maintaining safe and healthful environments and practices involves removal of hazards and provision of policies and procedures, as well as compliance with quality standards by everyone in the group.

Many of the directors said they appreciated the help they received from the CCHCs that ITQIP linked with their centers.
This selection bias is likely to have influenced the observed improvements.

A limitation of the study is the small sample size due to limited funding for the project. Also, although the study assessed practices for 13 CFOC3 standards (AAP et al., 2011), the centers addressed only three topic areas. Little improvement was seen in topics that were not chosen or chosen less frequently. Change in leadership at the centers with varying levels of interest in working on the action plans made improvement difficult.

Another limitation of the study is the variability in child care operation from one facility to another and from year to year. Evaluators were unlikely to have been evaluating the same children from pretest through Posttest 2. Different teachers/caregivers and children may occupy designated rooms in a facility. ITQIP did not require that the CCHCs spend a specific amount of time with their centers. The time and type of service provided by CCHCs varied widely. Although CCHCs reported the total time and types of services they provided, they were not asked to report the time spent in each type of service (onsite visits, phone calls, or e-mails).

CCHCs support health and safety practices and environments that prevent harm and promote health and development of children, as well as overall wellbeing for families and early education staff. Currently, only 17 U.S. states have a statutory requirement for early childhood education programs to have child care health consultation (Honigfeld, Pascoe, Macary, & Crowley, 2017). Of these, two states require CCHC involvement only if the facility cares for sick children (Honigfeld et al., 2017).

None of the centers in this project continued their relationship with their CCHC after the year of subsidized linkage. Some directors stated that although they found the CCHC very helpful and informative, the cost of the CCHC was prohibitive. Some said they would continue the CCHC on a fee basis if they could budget for it in the future. Other studies have shown that linkage of centers with CCHC improves health and safety compliance. ITQIP showed this is true for I/T programs, too.

REFERENCES


A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks

Richard Fiene, Ph.D.
Affiliated Faculty
Prevention Research Center
The Pennsylvania State University

This is a first of its kind study comparing the USA to other world countries utilizing the Child Care Aware – NACCRRA Child Care Benchmarks related to health and safety rules and regulations. A team of researchers analyzed the child care/early care & education rules and regulations from the USA and a selected group of countries to do a comparative analysis using the Child Care Aware – NACCRRA benchmarking scoring protocol. The results from the analyses were somewhat unexpected in that the scores between the USA and the other countries were not as statistically significant in the overall scores. However, when more specific benchmarks were compared statistically significant differences did appear in the health & safety and professional development areas.

Key words: Child Care Quality, Comparisons of USA and International Child Care, Child Care Regulations.

Introduction

The purpose of this paper is to compare several countries (N =20) and the United States on the Child Care Aware – formerly NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks that have used extensively in the USA to compare state regulatory and monitoring policy and implementation. The use of these benchmarks has been very useful in comparing states in the USA on an agreed upon series of child care benchmarks that have a great deal of support in the research literature (AAP/APHA, 2012, 2013; NACCRRA 2007, 2009, 2011). Previous research (OCED, 2006) has focused on early care and education policies in other countries which was a very important
first step in making comparisons across countries. This paper will expand upon this comparison in order to begin applying the NACCRRA benchmarks to other countries and establish a baseline between the USA and other countries related to regulatory review and analysis. This study is important because it provides a common rubric for making comparisons between the USA and other countries that is reliable and valid (NACCRRA 2007, 2009, 2011) related to regulatory analysis. As far as the author can determine from his extensive review of the literature, similar studies of this type have not been attempted utilizing a standardized rubric created by a major national child care organization. There have been other studies completed in which comparisons were made of other countries, the OCED (2006) Starting Strong II study and report is an excellent example of this type of

**Differential Monitoring Logic Model & Algorithm (DMLMA©) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model**

\[ \text{CI} \times \text{PQ} \Rightarrow \text{RA} + \text{KI} \Rightarrow \text{DM} \]

**Definitions of Key Elements:**

- \text{CI} = \text{Comprehensive Licensing Tool (Health and Safety) (Caring for Our Children)}
- \text{PQ} = \text{ECERS-R, FDCRS-R, CLASS, CDPES (Caregiver/Child Interactions/Classroom Environment)}
- \text{RA} = \text{Risk Assessment, (High Risk Rules) (Stepping Stones)}
- \text{KI} = \text{Key Indicators (Predictor Rules) (13 Key Indicators of Quality Child Care) (NACCRRA Benchmarks)}
- \text{DM} = \text{Differential Monitoring (How often to visit and what to review)}

![Diagram](image1.png)

Figure 1.
analysis and is recommended reading for anyone interested in reviewing public policy analyses.

The child care benchmarks utilized in this study are based upon the following key indicators: prevention of child abuse, immunizations, staff child ratio, group size, staff qualifications and training, supervision/discipline, fire drills, medication administration, emergency plan/contact, outdoor playground, inaccessibility of toxic substances, and proper hand washing/diapering (NACCRRA 2007, 2009, 2011). These benchmarks are more based upon the structural aspects of quality rather than on the process aspects of quality. This is an important distinction between the USA approach and the other countries approaches that becomes important in the explanation of results later in this paper.

This paper also supports and expands the development of an Early Childhood Program Quality Indicator Model (ECPQIM)(Fiene & Nixon, 1985) which is in a 4th generation (Fiene, 2013) as a differential monitoring logic model & algorithm helping to guide the program monitoring of child care/early care & education programs (see Figure 1).

**Method**

**Data Collection Process**

Data collection was done on a 100 point scale which is delineated in Appendix 1 as developed by the Child Care Aware - NACCRA Research Team. The same scoring protocol that was utilized in developing the 2007, 2009, and 2011 Reports and comparisons of states by Child Care Aware - NACCRA was employed in this study in comparing the average scores of the states and the 20 countries. The 100 point scale consisted of 10 child care benchmarks each worth 10 points: ACR = Staff child ratios NAEYC Accreditation Standards met (R1); GS = Group size NAEYC Accreditation Standards met (R2); Director = Directors have bachelor’s degree (R3); Teacher = Lead teacher has CDA or Associate degree (R4); Pre = Initial orientation training (R5); Inservice = 24 hours of ongoing training (R6); Clearance = Background check (R7); Devel = Six developmental domains (R8); Health = Health and safety recommendations (R9); and Parents = Parent Involvement (R10).

**Data Scoring**

The scoring protocol employed a total raw score approach of 100 points that was used to compare the countries on the 10 child care benchmarks in the aggregate. The scoring protocol also employed a standardized scoring approach (0 to 2 points) on each of the 10 child care benchmarks utilizing the following scale: 0.0 = Does not meet the Child Care Aware - NACCRA Benchmarks; 0.5 = Marginally meets the Child Care Aware - NACCRA Benchmarks; 1.0 = Partially meets the Child Care Aware - NACCRA Benchmarks; 1.5 = Substantially meets the Child Care Aware - NACCRA
Benchmarks; 2.0 = Fully meets the Child Care Aware – NACCRRA Benchmarks.

**Data Collectors**
A team of undergraduate and graduate research assistants at the Pennsylvania State University were the data collectors in which each of them reviewed the child care/early childhood rules/regulations/standards from a specific country and scored the rules/regulations/standards on the Child Care Aware – NACCRRA 100 point raw score protocol and the standardized (0 – 2) scoring approach.

**Data Sources**
The child care regulations selected were for preschool age children only in child care center setting in the 20 countries. Geographically the governmental jurisdiction closest to the national capital was used if applicable national regulations could not be found. More than the final 20 countries selected were reviewed but several countries needed to be dropped because they did not meet the above criteria or the regulations could not be found in English. This was more a convenience sample rather than a stratified scientific sample, a limitation of this study.

**Results**
The results from this study and analysis were totally unexpected. The results indicated no statistically significant differences between the USA and the other countries selected (Australia, Belgium, Norway, Finland, Sweden, Ireland, United Kingdom, Italy, France, New Zealand, Mexico, Greece, Canada, Austria, Portugal, Philippines, Turkey, Pakistan, Nigeria, Denmark, and Spain – these countries were selected because of their availability of child care/early care & education rules and regulations as described previously above in Data Sources) when comparing the total scores on the 100 point scale; the USA average for all 50 states scored 58 while the 20 countries average score was 56. However, a very different scenario occurs when looking at the ten individual child care benchmarks using the standardized 0 – 2 scoring protocol. The 20 countries selected in this study scored statistically higher on the following child care benchmarks: Director (t = 7.100; p < .0001) and Teacher (t = 7.632; p < .0001) qualifications. The USA scored statistically higher on the following child care benchmarks: Health/Safety (t = 6.157; p < .0001), Staff Clearances (t = 3.705; p < .01), and Pre-Service (t = 4.989; p < .001) /In-Service training (t = 2.534; p < .02) (See Table 1 & Figure 2).

The results showed that both the USA and all other countries mean scores were 58 and 56 respectively on the 100 point scale – this is a raw scale score and not the standardized score (0 – 2 – see Table 1 and Figure 2) which was used in the comparisons for each benchmark. This is not a particularly good score if you think in terms of exams, but for states and countries with
vastly complex bureaucracies maybe this isn’t as bad as it looks. Could it be that the USA is better than we think or is it that the USA and all other countries are providing just mediocre child care?!

The reason for using aggregate data in this study was to be consistent in how data have been collected in the USA utilizing the Child Care Aware – NACCRRA Scoring Protocol. This did delimit the potential analyses for this study and the recommendation would be made in future studies to unbundle the results so that more detailed comparisons could be made. As mentioned in the introduction, the purpose of this study was to provide an initial baseline comparison between the USA and other countries on the Child Care Aware – NACCRRA Scoring Protocol.

Table 1

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Countries</th>
<th>USA</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR (R1)</td>
<td>1.122</td>
<td>0.8462</td>
<td>not significant</td>
</tr>
<tr>
<td>GS (R2)</td>
<td>0.4063</td>
<td>0.5865</td>
<td>not significant</td>
</tr>
<tr>
<td>Director (R3)</td>
<td>1.5625</td>
<td>0.5</td>
<td>t = 7.100; p &lt; .001</td>
</tr>
<tr>
<td>Teacher (R4)</td>
<td>1.6563</td>
<td>0.4038</td>
<td>t = 7.632; p &lt; .001</td>
</tr>
<tr>
<td>Preservice (R5)</td>
<td>0.9375</td>
<td>1.6731</td>
<td>t = 4.989; p &lt; .001</td>
</tr>
<tr>
<td>Inservice (R6)</td>
<td>0.6563</td>
<td>1.0481</td>
<td>t = 2.534; p &lt; .02</td>
</tr>
<tr>
<td>Clearances (R7)</td>
<td>0.6094</td>
<td>1.2404</td>
<td>t = 3.705; p &lt; .01</td>
</tr>
<tr>
<td>Development (R8)</td>
<td>1.6406</td>
<td>1.4519</td>
<td>not significant</td>
</tr>
<tr>
<td>Health (R9)</td>
<td>0.9844</td>
<td>1.7404</td>
<td>t = 6.157; p &lt; .0001</td>
</tr>
<tr>
<td>Parent (R10)</td>
<td>1.5000</td>
<td>1.5385</td>
<td>not significant</td>
</tr>
</tbody>
</table>

Legend:
Child Care Aware - NACCRRA Benchmarks:
Parent = Parent Involvement (R10)
Health = Health and safety recommendations (R9)
Development = Six developmental domains (R8)
Clearances = Background check (R7)
Inservice = 24 hours of ongoing training (R6)
Preservice = Initial orientation training (R5)
Teacher = Lead teacher has CDA or Associate degree (R4)
Director = Directors have bachelor’s degree (R3)
GS = Group size NAEYC Accreditation Standards met (R2)
ACR = Staff child ratios NAEYC Accreditation Standards met (R1)

Scoring:
0.0 = Does not meet Child Care Aware – NACCRRA Benchmarks.
0.5 = Marginally meets Child Care Aware – NACCRRA Benchmarks.
1.0 = Partially meets Child Care Aware – NACCRRA Benchmarks.
1.5 = Substantially meets Child Care Aware – NACCRRA Benchmarks.
2.0 = Fully meets Child Care Aware – NACCRRA Benchmarks.
Discussion

The purpose of this study was to extend the Child Care Aware - NACCRRRA Child Care Benchmarks Scoring Protocol to an international sample comparison. As has been done by the National Science Foundation with math and science testing, these same types of comparisons have been made with the USA not fairing all that well on the math and science comparisons.

It appears that when it comes to child care benchmarks the USA actually appears to be in better shape than many advocates and experts would have thought when compared to other countries or is it that the other countries are providing the same form of mediocre care as it relates to these child care benchmarks. Remember that these benchmarks are heavily weighted towards the structural side of quality

Legend:
Child Care Aware - NACCRRRA Benchmarks:
Parents = Parent Involvement (R10)
Health = Health and safety recommendations (R9)
Devel = Six developmental domains (R8)
Clearance = Background check (R7)
Inservice = 24 hours of ongoing training (R6)
Pre = Initial orientation training (R5)
Teacher = Lead teacher has CDA or Associate degree (R4)
Director = Directors have bachelor’s degree (R3)
GS = Group size (R2)
ACR = Staff child ratios (R1)

Scoring:
0.0 = Does not meet Child Care Aware – NACCRRRA Benchmarks.
0.5 = Marginally meets Child Care Aware – NACCRRRA Benchmarks.
1.0 = Partially meets Child Care Aware – NACCRRRA Benchmarks.
1.5 = Substantially meets Child Care Aware – NACCRRRA Benchmarks.
2.0 = Fully meets Child Care Aware – NACCRRRA Benchmarks.

Figure 1. Mean Comparisons between USA and Twenty Countries on Child Care Aware – NACCRRRA Benchmarks
rather than the process side of quality.

However, when the individual benchmarks are analyzed then certain patterns occur which seem very consistent with the previous research literature. The 20 countries scored higher on the staffing benchmarks while the USA scored higher on the training and health/safety benchmarks. Clearly this is an indication reflecting public policy in the other countries as versus the USA. Many other countries place more emphasis on the process aspects of quality which involve staff and staff interactions with children. The USA has focused more on the structural aspects of quality which involve health & safety especially in the state licensing of child care. These structural aspects of quality are more easily quantifiable in state rules and regulations which is the locus of control for the licensing of child care. Since the USA does not have national standards that are required (the USA does have national health and safety standards that are recommended practice, such as Caring for Our Children (2012)) as is the case in so many of the countries in this study, this may provide a possible explanation for the results of this study. It will be interesting to see how Quality Rating and Improvement Systems (QRIS) which usually have some process standards impact this overall balance of structural and process aspects of quality. This is an area that needs additional research and more in-depth analysis.

So what does this tell us. I think it is a warning call as has been put forth by Child Care Aware - NACCRA that we still have a lot of additional work to do in improving child care, not only in the USA, but worldwide. Just as the Child Care Aware - NACCRA Report Cards (2007, 2009, 2011) have played a role in making positive change in the child care benchmarks over time; we need to expand this reporting and change to a world wide focus. There is clearly the need to expand from the present analysis of 20 countries and the USA to other countries throughout the world and to track changes over time as Child Care Aware/NACCRA has done.

Another area of concern within the USA and I am sure in other countries as economies have begun their slow recovery from the economic downturn of 2008 – 2010 is to do more with less. One such approach being explored in the USA is called differential monitoring which helps to re-allocate limited resources in a more cost effective and efficient manner via a risk assessment and key indicator approach. I hope that this comparison utilizing the Child Care Aware – NACCRA Benchmarking Scoring Protocol and introducing the Early Childhood Program Quality Indicator Model/Differential Monitoring Logic Model and Algorithm (Fiene, 2013) within an international context as first steps in making that happen.
References


Notes

1 In the licensing literature these child care benchmarks are usually referred to as key indicators (Fiene, 2013). Please see Figure 1 which delineates where within a program monitoring system these benchmarks would appear and could be utilized.

2 The following individuals played key data collection roles as research assistants in the compilation of this study: Melissa Cave, Ashley Le, Breanna Green, Corrie Podschln, Sherrie Laporta, Ashley Edwards, Laura Hartranft, Gissell Reyes, Janet Lazur, Kayma Freeman, Jessica White, Karen Mapp, and Lindsay Bitler.
Appendix 1

Benchmark criteria for *We Can Do Better: NACCRA Ranking of State Child Care Center Regulations: 2011 Update* were developed by Child Care Aware - NACCRA and have been used for the 2007, 2009 and 2011 *We Can Do Better* reports. The rationale for each standard, including research evidence of its importance in quality care, is noted in each section of the report and in previous reports. Each of the 10 regulation benchmarks were scored with a value ranging from one to 10 points, depending on how closely the state met the benchmark, for a maximum total of 100 points. In cases where states permit several different options for complying (e.g., complying with director or teacher qualifications), the minimum allowed was used. This information was used to generate state sheets with scores for each standard.

### Scoring Methods for NACCRA Ranking of State Child Care Center Regulations (R)

<table>
<thead>
<tr>
<th>Question</th>
<th>Scoring method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regulation 1.</strong> Staff:child ratio requirements comply with NAEYC accreditation standards.</td>
<td><strong>Number of ratios in compliance with NAEYC standards</strong></td>
</tr>
<tr>
<td>6 mo</td>
<td>9 mo</td>
</tr>
<tr>
<td>1:4</td>
<td>1:4</td>
</tr>
<tr>
<td>7 ratios</td>
<td>10</td>
</tr>
</tbody>
</table>

| **R2.** Group size requirements are in compliance with NAEYC accreditation standards. | **Number of group sizes in compliance with NAEYC standards** | **Score** |
| 6 mo | 9 mo | 18 mo | 27 mo | 3 yr | 4 yr | 5 yr | |
| 8 | 8 | 8 | 8 | 18 | 20 | 20 | |
| 7 ratios | 10 | 6 ratios | 9 | 5 ratios | 8 | 4 ratios | 7 | 3 ratios | 5 | 2 ratios | 3 | 1 ratios | 1 |
R3. Center directors are required to have a bachelor’s degree of higher in early childhood education or a related field.

<table>
<thead>
<tr>
<th>Director education requirement</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor’s degree in any field</td>
<td>10</td>
</tr>
<tr>
<td>College directors certification</td>
<td>7</td>
</tr>
<tr>
<td>Any associate degree</td>
<td>5</td>
</tr>
<tr>
<td>CDA</td>
<td>5</td>
</tr>
<tr>
<td>Clock hours/less than associate degree</td>
<td>2</td>
</tr>
<tr>
<td>High school or less</td>
<td>0</td>
</tr>
</tbody>
</table>

R4. Lead teachers are required to have a Child Development Associate (CDA) credential or an associate degree in early childhood education or related field.

<table>
<thead>
<tr>
<th>Lead teacher education requirement</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDA/associate degree or better</td>
<td>10</td>
</tr>
<tr>
<td>State Credential</td>
<td>5</td>
</tr>
<tr>
<td>Clock Hours in ECE</td>
<td>2</td>
</tr>
<tr>
<td>High School/GED</td>
<td>2</td>
</tr>
<tr>
<td>Less than High School</td>
<td>0</td>
</tr>
</tbody>
</table>

R5. Lead teachers are required to have initial training, including:
- Orientation.
- Fire safety.
- Other health and safety issues.
- At least one staff member certified in first aid must be present when children are in care.
- At least one staff member who is certified in CPR must be present when children are in care.

<table>
<thead>
<tr>
<th>Number of areas training is required</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five areas</td>
<td>10</td>
</tr>
<tr>
<td>Four areas</td>
<td>8</td>
</tr>
<tr>
<td>Three areas</td>
<td>6</td>
</tr>
<tr>
<td>Two areas</td>
<td>4</td>
</tr>
<tr>
<td>One area</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
</tbody>
</table>

R6. Lead teachers are required to have 24 hours or more of annual training.

<table>
<thead>
<tr>
<th>Ongoing training &gt;</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 Hours</td>
<td>10</td>
</tr>
<tr>
<td>18 hours</td>
<td>7</td>
</tr>
<tr>
<td>12 hours</td>
<td>5</td>
</tr>
<tr>
<td>6 hours</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
</tbody>
</table>

R7. A comprehensive background check is required for child care providers.
- Use of fingerprints to check state records.
- Check FBI records.
- Check state child abuse registry
- Check sex offender registry.
- Criminal history check.

<table>
<thead>
<tr>
<th>Number of Background checks completed</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five checks</td>
<td>10</td>
</tr>
<tr>
<td>Four checks</td>
<td>8</td>
</tr>
<tr>
<td>Three checks</td>
<td>6</td>
</tr>
<tr>
<td>Two checks</td>
<td>4</td>
</tr>
<tr>
<td>One check</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
</tbody>
</table>
A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA Child Care Benchmarks

Appendix 2

These were the countries included in these analyses: Australia, Belgium, Norway, Finland, Sweden, Ireland, United Kingdom, Italy, France, New Zealand, Mexico, Greece, Canada, Austria, Portugal, Philippines, Turkey, Pakistan, Nigeria, Denmark, Spain, and the USA which included all 50 states.
Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs

Michel Lahti a,*, James Elicker b, Gail Zellman c, Richard Fiene a,b,c,1

a University of Southern Maine, Muskie School of Public Service, PO Box 9300, Portland, ME 04104-9300, United States
b Purdue University, United States
c RAND Corporation, United States

ARTICLE INFO
Article history:
Available online xxx

Keywords:
Child care quality
QRIS
Study design
Methodology

Abstract
In recent years, child care quality rating and improvement systems (QRISs) have become an increasingly popular policy tool to improve quality in early childhood education and care (ECEC) settings and have been adopted in many localities and states. The QRIS proposition is that with higher-quality child care settings, it is more likely that children who attend those high-quality programs will benefit in terms of outcomes like school readiness. However, in order to demonstrate this linkage, QRIS standards and ratings must function as intended, i.e. be valid. This paper presents a framework for validating child care quality improvement standards and processes, along with examples from recent QRIS validation studies in two states. The state examples provide useful data about the strengths and limitations of these validation approaches. We discuss the implications of applying these approaches and provide recommendations to researchers, policy-makers, and program leaders who implement QRIS validation studies.

© 2014 Elsevier Inc. All rights reserved.

Introduction

In recent years, child care quality rating and improvement systems (QRISs) have become an increasingly popular policy tool to improve quality in early childhood education and care (ECEC) settings and have been adopted in many localities and states. The QRIS National Learning Network reports that 40 statewide QRISs have launched or piloted, including the District of Columbia (QRIS National Learning Network, 2014). The immediate goal of a QRIS is to raise the quality of care in early learning settings. Existing research suggests that care in higher-quality settings will improve child functioning, including school readiness (Burchinal et al., 2009; Burger, 2010; Howes et al., 2008), especially for children from lower-income families. QRIS logic models that guide these large-scale interventions focus on improving various dimensions of ECEC quality, with the ultimate goal of improving system outcomes, namely, child care program quality, training and technical assistance for child care providers, information and support for families, and, therefore, improvements to children’s cognitive, language, social, emotional, and physical development.

The perceived need for QRIS has grown out of documented gaps in quality in existing ECEC programs, especially those serving children from lower-income families (Fuller, Loeb, Kagan, & Carrol, 2004; NICHD ECCRN, 2000) and the inability of the current ECEC system to promote uniformly high quality (Cochran, 2007). QRISs produce program-level quality ratings based on multi-component assessments designed to make ECEC quality transparent and easily understood to parents and other stakeholders. Most also include feedback, technical assistance, and incentives to both motivate and support providers’ efforts toward quality improvement (Tout et al., 2010). To make program quality transparent, QRISs typically rely on a multi-tiered rating system with one to five levels of program quality. Therefore, it is important that these ratings show evidence of validity, so that higher-quality programs are rated higher, and lower-quality programs are rated lower.

Recent research has documented the importance of both specificity and thresholds when testing hypotheses about child care quality impacts on children’s developmental outcomes (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Howes, Whitebook, & Phillips, 1992; NICHD ECCRN, 2000, 2002). However, common global measures of classroom quality such as the Early Childhood Environment...
Rating Scale- Revised (ECERS-R; Harms, Clifford, & Cryer, 2005) are not always significantly associated with specific child outcomes (Burchinal, Kainz, & Cat, 2011). This may be because these global quality scales do not focus enough on the particular child care quality processes most likely to bring about improved child outcomes (specificity) or they do not provide guidance for the level of quality required to produce improved child outcomes (thresholds). As states implement QRISs, they are using observational measures such as the ECERS-R, and they may also combine other quality measures such as the Classroom Assessment Scoring System (CLASS: Pianta, La Paro, & Hamre, 2008) or locally specified quality indicators. Because QRIS quality standards are often complex, including many components and measures at several quality levels, and because they vary from state to state, it is especially important for states to carefully validate their quality rating systems and match measures specifically to the stated outcome goals of the QRIS. For example, if a particular QRIS places more emphasis on the health aspects of children’s development, then the ECERS-R and CLASS would not be appropriate tools; but a tool measuring child care health indicators, such as the National Health and Safety Tool being developed by the California Child Care Health Program (Alkon, 2013) would be more appropriate.

Validity data can also enable researchers to test conclusions about whether the quality indicators embedded in QRIS standards lead to adequate quality assessment and whether the methods used to assign quality ratings are working as intended (Cizcek, 2007). This paper defines operationally the concept of QRIS validity, presents four general approaches to assessing validity in the context of large-scale QRISs, and critically examines the efforts of two states, Maine and Indiana, to assess the validity of recently implemented QRISs using these approaches.

Validation of a QRIS is a developmental and multi-step process that assesses the degree to which decision designs about program quality standards and measurement strategies are resulting in accurate and meaningful quality ratings. Validation of a QRIS provides designers, administrators, and stakeholders with crucial data about how well the system is functioning. A carefully designed plan for ongoing QRIS validation creates confidence in the system and a climate that supports continuous quality improvement at both the child care provider and system levels (Zellman & Fiene, 2012).

To date, QRIS validation research efforts have been limited, for a number of reasons. First, validation is complex and involves a range of activities, which should include validating standards, measures, and rating protocols. Second, there has been little information available in the field that clarifies the importance and purpose of QRIS validation or identifies recommended strategies. Third, child care quality advocates and policy makers have been extremely busy designing and implementing these statewide systems, often with limited resources. Given these constraints, validation may seem like an abstract luxury that can wait until later. Further, in states with more mature QRISs, there may be some reluctance among stakeholders to assess the validity of an established and accepted quality improvement system. In newer state systems, policymakers may question the need for validation, given arguments recently offered in support of establishing a QRIS system (Zellman & Fiene, 2012; Zellman, Brandon, Boller, & Kreader, 2011). Yet early and ongoing validation research is essential to the long term success of any system.

One challenge is that QRIS validation cannot be determined by a single study. Instead, validation should be viewed as an iterative process with several equally important goals: refining the QRIS quality standards and ratings, improving system functioning, and increasing the credibility and value of rating outcomes and the QRIS system as a whole. A carefully designed validation plan can promote the accumulation of evidence over time that will provide a sound theoretical and empirical basis for the QRIS (AERA, APA, & NCME, 1999; Kane, 2001; Zellman & Fiene, 2012). Ongoing validation activities, carried out in tandem with QRIS monitoring activities (those that examine ongoing implementation processes) and evaluation activities (those that examine specific outcomes) can help a QRIS improve throughout its development, implementation, and maturation (Lugo-Gil et al., 2011; Zellman et al., 2011).

QRIS validation research may produce three important benefits. First, validation evidence can promote increased support for the system among parents, ECCE providers, and other key stakeholders. Ratings that mirror the experiences of parents and providers can build trust and increase the overall credibility of the system. Second, a system that is measuring quality accurately and specifically should better able to target limited quality improvement resources to programs and program elements most in need of improvement. This should result in more targeted and effective supports for programs striving to offer higher-quality services. Third, validation evidence can be used to improve the efficiency of the rating process. If a QRIS is expending resources to measure a component of quality that is not making a unique contribution to a summary quality rating, is not measuring quality accurately, or is not contributing to desired program outcomes, that component can be removed or revised. For example, measures that vary little across providers, whose quality varies substantially in other ways, make little or no contribution to overall quality ratings (Zellman & Fiene, 2012).

Four approaches to validation

A comprehensive QRIS validation plan includes multiple studies that rely on different sources of information and ask different but related questions. We suggest QRIS validation research be organized around four complementary approaches: key quality concepts; quality measurement; ratings outputs; and links to child outcomes (Zellman & Fiene, 2012). Summaries of these approaches are provided in Table 1, which includes the purpose of each validation approach, the types of research that can be undertaken, the questions that are asked, and some limitations of each approach. The four approaches are also elaborated later in the paper, as we summarize results of validation research in Indiana and Maine.

In reviewing the table, and throughout this paper, we use three key QRIS terms: component, standard, and indicator. The term ‘quality component’ refers to broad quality categories used in QRIS (such as staff qualifications, family engagement, or learning environment). A ‘quality standard’ is defined as a specific feature of quality, such as specialized training in the use of developmentally appropriate curriculum or developmental assessment training within the staff qualifications component. A set of quality standards comprise each quality component. ‘Quality indicators’ are the specific metrics used for each quality standard. A given quality standard may have one or more quality indicators. An indicator related to the curriculum/assessment staff training standard may be, for example, “At least 50% of teaching staff have completed the two-course statewide training session on developmentally-appropriate curriculum.”

QRIS validation in Indiana and Maine

This section will describe efforts at QRIS validation in two states in order to explore current validation efforts using these four approaches and to identify the successes and challenges experienced in these early QRIS validation studies. In Indiana and Maine, the QRIS designs are similar, but some aspects of the states’ child care contexts, specific QRIS quality components, standards, and rating processes employed are somewhat different. Both states launched their QRIS statewide in 2008, and both systems have four quality tiers, referred to as “levels” in Indiana and “steps” in Maine, organized into a “building block” framework, meaning that child

Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. Early Childhood Research Quarterly (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005
Table 1
Four related approaches to validating a QRIS.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Activities and purpose</th>
<th>Typical questions</th>
<th>Issue and limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examine the validity of key underlying concepts.</td>
<td>Assess whether basic QRIS quality components and standards are the &quot;right&quot; ones to include by examining levels of empirical and expert support.</td>
<td>*Do the quality components capture the key elements of quality?</td>
<td>*Process subject to interpretation and to political pressure</td>
</tr>
<tr>
<td>2. Examine the measurement strategy and psychometric properties of measures used to assess quality.</td>
<td>*Examine properties of key quality measures, e.g., inter-rater reliability on observational measures, scoring of documentation, and inter-item correlations, to determine if measures are psychologically sound.</td>
<td>*Is there sufficient empirical and expert support for including each standard?</td>
<td>*Limited empirical evidence available; few established links to outcomes of interest.</td>
</tr>
<tr>
<td></td>
<td>*Examine relationships among quality measures to assess whether they function as expected.</td>
<td>*What is the reliability and accuracy of indicators collected using different methods?</td>
<td>*This validation activity is especially important given that some quality component measures were likely developed in low-stakes settings and have not been examined in the high-stakes context of QRIS.</td>
</tr>
<tr>
<td>3. Assess the outputs of the rating process</td>
<td>*Examine variation and patterns of program-level ratings within and across program types, to assess if QRIS distinguishes levels of quality.</td>
<td>*Do quality measures perform as expected? (e.g., do expected subscales emerge?)</td>
<td>*Measurement error is an important issue that should be examined.</td>
</tr>
<tr>
<td></td>
<td>*Examine relationship of program-level ratings to other validated quality indicators to determine if ratings are assessing quality in expected ways.</td>
<td>*Do measures of similar standards relate more closely to each other than to other measures?</td>
<td>*These validation activities depend on a reasonable level of confidence about the quality components, standards and indicators as well as the process used to designate ratings.</td>
</tr>
<tr>
<td></td>
<td>*Examine alternate cut points and combining rules to determine how well the ratings distinguish different levels of quality.</td>
<td>*Do measures relate to each other in theoretically consistent ways?</td>
<td>*Comparing QRIS measures to other measures is frequently constrained by the absence of validated alternative measures of the same constructs.</td>
</tr>
<tr>
<td>4. Examine how ratings are associated with children's outcomes.</td>
<td>Examine the relationship between program-level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes.</td>
<td>*Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?</td>
<td>*Appropriate demographic and program level control variables must be included in analyses to account for selection factors.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Do levels cut scores and combining rules produce expected rating distributions and meaningful distinctions among programs?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do children who are attending higher-rated programs have greater levels or gains in skills than children who attend lower-quality programs?</td>
<td>*Studies may be done on selected child and program samples to save resources.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Do children who are attending higher-rated programs have greater levels or gains in skills than children who attend lower-quality programs?</td>
<td>*Given staff turnover and movement, children may not spend substantial time with quality-rated caregivers (dosage.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Are children who are attending higher-rated programs more likely to be retained in care?</td>
<td>*Measurement limitations noted for Approach 3 apply here as well.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do children who are attending higher-rated programs have greater levels or gains in skills than children who attend lower-quality programs?</td>
<td>*Findings from non-experimental studies do not permit attribution of causality about QRIS participation, but preliminary inferences can be made about how quality influences children’s outcomes.</td>
</tr>
</tbody>
</table>

care providers must enter at the lowest level and meet all quality standards and indicators at each level in order to advance to the next higher level. The focus on these two states in this paper is to help illustrate the application of these four approaches to operationalizing validation in a QRIS. While the QRIS evaluations in Maine and Indiana have resulted in other kinds of information disseminated for policy makers in these states and publications for other audiences, this paper is unique in that it is only intended to focus on these four concepts of validation.

Both states partnered with university-based researchers to conduct validation research, after piloting aspects of their QRIS design. However, there are also key differences between these two states. For example, the Indiana QRIS standards were developed based on a local community-based model that was then modified by a state stakeholder committee for statewide expansion. The Maine quality standards were developed to align with program-type-specific national accreditation standards. The Maine standards were also vetted through review and comment by many stakeholders and technical assistance was provided by University researchers based on reviews of the scientific literature. Maine QRIS ratings are generated by provider self-report, then verified by state agency staff, while Indiana employs independent raters who directly assess the standards by visiting child care settings. Provider voluntary participation rates are higher among state-licensed providers in Indiana. However, Indiana also has significant numbers of license-exempt child care providers, whereas license exemption is not a prominent feature of the Maine child care system. The key features of each state QRIS are summarized in Table 2. These two states provide useful examples, because while the state child care contexts are different, they each used strategies contained in the four validation approaches discussed above and outlined in Table 1. The successes and limitations of these states’ approaches will inform future validation research on QRIS.
Table 2
Key features of Indiana and Maine QRISs.

<table>
<thead>
<tr>
<th>QRIS feature</th>
<th>Indiana</th>
<th>Maine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible child care types (participation rate)</td>
<td>Licensed centers (89%)</td>
<td>Licensed centers (68%)</td>
</tr>
<tr>
<td></td>
<td>Licensed homes (62%)</td>
<td>Licensed homes (44%)</td>
</tr>
<tr>
<td></td>
<td>Unlicensed registered ministry centers (12%)</td>
<td></td>
</tr>
<tr>
<td>Participation rules</td>
<td>Voluntary for all providers (Not required for CCDF participation)</td>
<td>Required for programs participating in federal</td>
</tr>
<tr>
<td></td>
<td>Building Block; 4 quality levels</td>
<td>Building Block; 4 quality levels</td>
</tr>
<tr>
<td>QRIS structure</td>
<td>Level 1 – Licensed, or completes voluntary certification program</td>
<td>Step 1 – Meets all regulatory standards, in operation for more than one year, and all staff registered in Maine Roads to Quality Registry (MRTQ);</td>
</tr>
<tr>
<td>QRIS standards (examples)</td>
<td>Level 2 – Learning environment and materials requirements; daily literacy activities; 25% of staff have CDA or equivalent; 15 hrs. in-service training/yr.; etc.</td>
<td>Step 2 – Learning Environment/Developmentally Appropriate Practice requirements; program improvement plan in place; 50% of staff at level 5 on MRTQ career lattice; etc.</td>
</tr>
<tr>
<td></td>
<td>Level 3 – Written curriculum focused on whole child; provision for special needs; 50% of staff have CDA or equivalent; 20 hrs. in-service training/yr.; etc.</td>
<td>Step 3 – Documented use of Early Childhood Learning Guidelines and/or Infant-Toddler Learning Guidelines; Evidence collected at least three times per year on child’s development; etc.</td>
</tr>
<tr>
<td></td>
<td>Level 4 – National accreditation; Provide mentoring to other QRIS providers (see <a href="http://www.in.gov/fssa/carefinder/2554.htm">www.in.gov/fssa/carefinder/2554.htm</a>).</td>
<td>Step 4 – National accreditation; written parent involvement plan; etc. (See: <a href="https://www.maine.gov/dhhs/oecs/ecs/qrisc/qualityforme.htm">https://www.maine.gov/dhhs/oecs/ecs/qrisc/qualityforme.htm</a>).</td>
</tr>
<tr>
<td>QRIS standards development process</td>
<td>Aligned with national accreditation standards; expansion of community pilot program, modified by state stakeholder committee.</td>
<td>Aligned with national accreditation standards; based on available empirical evidence; and from results of stakeholder reviews and comment.</td>
</tr>
<tr>
<td></td>
<td>(1) Provider self-assessment, online system; (2) enrollment system uses linked files from state licensing and registry; and (3) verified by state agency staff.</td>
<td></td>
</tr>
</tbody>
</table>

Method

Indiana

The Indiana QRIS is called “Paths to QUALITY™.” The validation research reported here includes a preliminary literature review and an empirical field study including a stratified random sample of 276 child care providers who had voluntarily entered the QRIS during 2008–2009, including 135 classrooms in 95 licensed child care centers, 169 licensed family child care homes, and 14 classrooms in 12 unlicensed registered child care ministry centers. Independent, on-site assessments were completed by university researchers approximately one year after QRIS entry and included: observational global quality assessments of the child care environment using the Environmental Rating Scales (ERS; ITERS-R, ECERS-R, FCCERS-R; Harms et al., 2005); observations of adult–child interaction quality (Caregiver Interaction Scale, CIS; Arnett, 1989); surveys and interviews with child care providers; and interviews with parents whose children had been placed with QRIS providers. Observers were trained to reliability level of 80% exact agreement (Kappa = .70) or higher, and maintained reliability during the study. Child development assessments were completed using standardized research–validated measures, with two randomly selected children from each participating child care center classroom or family child care home. For children under three years, measures included the Mullen Scales of Early Learning (Mullen, 1995) for cognitive and language development and the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan & Carter, 2002) for social–emotional development. For children three to five years, the measures included the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 1997) and Woodcock-Johnson Applied Problems and Letter–Word Identification subtests (Berry, Bridges, & Zaslav, 2004) for language and cognitive development and the Social Competence and Behavior Evaluation (SCBE; LaFreniere & Dumas, 1997) for social–emotional development. (For a detailed description of the Indiana evaluation methodology, see Elicker et al., 2013; Elicker, Langill, Ruprecht, Lewsader, & Anderson, 2011.)

Maine

The Maine QRIS is called “Quality for ME.” The Maine validation research reported in this paper is based on a literature review of quality variables, focus group interviews with providers and parents, and a field study including a stratified random sample of 255 providers who enrolled in the QRIS in 2008 through 2011, including: 153 classrooms in 105 licensed child care centers; 113 licensed family child care homes; and 41 classrooms in 37 Head Start sites. Assessments were completed as soon as possible after a program enrolled into the QRIS, however, this varied based on the length of time required for the state agency to verify enrollment information and schedule on-site observations. Like Indiana, Maine researchers used the ERS global quality assessment scales and conducted surveys with providers and parents. Unlike Indiana, Maine did not collect any child-level outcome data. Assessors were trained to reliability annually by authors of the ERS scale and maintained a 85% inter-rater reliability during the study. (For a detailed description of the evaluation methodology, see Lahti et al., 2011.)

Results

Results of the QRIS validation research in Indiana and Maine are presented in relation to the four approaches to validation recommended by Zellman and Fienie (2012; refer to Table 1).

Approach 1: examine the validity of key underlying concepts

As noted above, the quality components included in a QRIS (e.g., staff qualifications, learning environment, family engagement) essentially define how child care quality will be viewed in each state. Conceptual validation provides justification and support for these chosen elements. This first validation approach asks whether the quality components, standards, and indicators included in a QRIS are the “right” ones; that is, if together they define quality of care. Many state QRISs have adopted similar, though not identical, concepts and program quality standards (Smith, Robbins, Stagman, & Kreader, 2012).
One approach that can help to validate the underlying concepts of quality in a QRIS involves assessing the degree to which the quality components used in the QRIS ratings include standards and indicators that are based on empirical evidence that links them to desired program, family and child outcomes. A literature review weighs the existing research evidence and on that basis provides a judgment about whether a particular quality component should be included or excluded from the QRIS. Like many validation activities, such reviews ideally would be updated from time to time to determine if revisions to the QRIS are advisable in light of new research findings. As noted in Table 1, this approach may be limited by available data. Further, available data may be subject to more than one interpretation. Politics can also play a role: supporters of particular elements, e.g., nutrition, accreditation, may want to ensure that such measures are included, regardless of the strength of the research evidence. This literature review approach of conceptual validation was a key method used in developing both Indiana’s and Maine’s QRIS quality standards.

Indiana: examining the validity of underlying concepts

Standards and indicators for each QRIS level in Indiana were drafted by a state committee of child care providers and stakeholders. The standards were based on an existing community-level Paths to QUALITY model, but also made accommodations for statewide use and integration into the existing state child care licensing and training/technical assistance systems. The highest level quality goal for QRIS in Indiana is national accreditation, so proposed quality standards and criteria at each QRIS level were constructed to help child care providers work toward accreditation in steps.

The Purdue University research team conducted a review of previous evaluations of the Indiana QRIS community-level pilot programs and an in-depth analysis of the proposed QRIS quality standards based on the published child development and child care literature. This literature-based analysis is summarized here. (For a full report, see Elicker, Langill, Ruprecht, & Kwon, 2007; Elicker et al., 2013.)

First, the evaluators looked at each proposed QRIS quality standard and indicator for each type of child care. Ten broad quality components were identified that encompassed all of the proposed quality indicators: regulation; teacher education/training; structural/environmental quality; process quality/interactions; assessment; provisions for children with special needs; program policies; director/owner professional development; parent–teacher communication; and national accreditation. These ten components were then used as key terms to guide an extensive search of the research literature to collect and weigh the available evidence that each component was: (1) generally considered a valid aspect of quality; and (2) empirically associated with children’s well-being or positive developmental outcomes. Based on the amount and quality of evidence, each quality component was the rated as follows: (1) some or limited evidence (one or two well-designed studies); (2) moderate evidence (3–5 well-designed studies); or (3) substantial evidence (more than five well-designed studies). The results of this analysis were reported to the state QRIS planning committee, including a conclusion that most of the proposed quality indicators had “substantial evidence” for their validity.

Maine: examining the validity of underlying concepts

Researchers at the University of Southern Maine worked with state agency leaders and other key stakeholders through a process that involved the use of Concept Mapping (The Concept System®, 2012). This process allows for the development of a conceptual framework that can guide planning, and in this case led to the selection of the underlying quality concepts and standards for Maine’s QRIS. Similar to what was done in Indiana, University of Southern Maine research staff identified key quality concepts from the literature and national accreditation standards. In addition, concepts emerged from results of eight focus groups with parents and ECEC professionals across the state, including participants from various types of settings, e.g., family child care homes, after school programs, centers, etc. Statements of program quality were developed; these statements were the focus of a mapping process which involved more than twenty-four experts reviewing and rating the statements. The Concept Maps that result from this process allowed participants to visually identify which concepts of program quality were most favored by specific key stakeholder groups. In addition, the mapping software illustrated how closely related the concepts were to each other, based on reviews from the select experts. From this process, a set of components and standards was developed. The final step in selecting program quality standards involved a formal review and comment process that the state agency implemented in various locations across the state (Maine DHHS, 2008).

Approach 2: examine measurement strategies and psychometric properties of quality measures

A second type of validation focuses on the attributes of the individual quality measures used in the QRIS and the way these measures are combined to produce a summary rating of program quality. This approach addresses how well measures are working in the context of the QRIS. These efforts attempt to answer questions such as, “Is there evidence that a given indicator measures what it purports to measure?” “If the QRIS claims to have a specific number of dimensions, do we find those dimensions in the output data?” “Is there sufficient variance in scores on this quality indicator to justify its inclusion in the QRIS?” Addressing these issues involves an examination of the distribution of participating provider quality scores and the internal consistency of multi-item measures.

The research literature provides limited guidance concerning the most appropriate ways to combine measures of quality indicators into summary ratings (Lugo-Gil et al., 2011; Tout, Zaslow, Halle, & Forry, 2009; Zellman, Perlman, Le, & Setodji, 2008). Yet this process is crucial to producing meaningful overall program quality ratings, the key output of the rating assessment process. At minimum, it is important to consider whether certain elements should be treated as more important, and if so, how this can be assured in the process of combining them. If this issue is not addressed, unexamined weighting may occur anyway. For example, if measures of individual quality elements are combined without any weighting, then those measures that are longer (e.g., include more items) will count for more in a final rating.

At the time both the Maine and Indiana QRISs were being designed, in the mid-2000s, the predominant global quality measures in use in both states were the Environmental Rating Scales (ERS) (Harms et al., 2005; Harms, Cryer, & Clifford, 2006; Harms, Cryer, & Clifford, 2007). ERS use was predominant in the accreditation quality improvement efforts in both child care centers and child care homes. So there was some familiarity with the measures on the part of providers. This was an important political consideration in terms of developing and promoting the design of the QRIS. In addition, in reviewing emerging QRIS work from other states, it appeared that the ERSs were the predominant global classroom quality measure in use at that time. While ERS was influential in the design of the QRIS quality standards in both Maine and Indiana, it is important to note that the ERS are not used to determine the step or level quality ratings. Many other quality indicators are included in the QRIS standards of both states, including staff qualifications, annual staff training hours, and other indicators that help providers make progress toward the ultimate quality goal of national accreditation.
Choosing the points at which individual measures (in block design QRISs) and summary ratings are assigned to rating levels is another exercise that has received limited attention. Cut scores can be assessed in a number of ways. One relatively simple one is to use existing data to conduct a “virtual pilot” (Zellman & Karoly, 2012a) in which existing data are used and cut scores are altered and the effects are examined in terms of distributions of summary ratings across programs. A downward limit on cut scores is the need for some variation within each quality component; without it, a component provides no useful information in overall ratings. Designers may compare program distributions using different cut scores, although it is not always clear what an appropriate rating levels distribution should be. However, it is reasonable to assume that an appropriate distribution in the early phase of a QRIS would be one in which there are programs placed at all levels, with decreasing numbers of programs at each succeeding higher level.

Another validation activity might involve an assessment of the relationship of a given indicator to other indicators of quality included in the QRIS. In studies that examine measures to be included together in a QRIS, it is important to look at the degree of correlation found among these measures: ideally, measures will be moderately correlated so that each measure both contributes to an overall assessment of quality yet also provides some non-redundant program quality information (Zellman et al., 2008). Correlation patterns should make sense. For example, two measures of interaction quality should be more closely related to each other than to a measure of adult–child ratios. If such studies reveal for example that the correlation between ratios and interaction processes is very high (r = .90+) this result might argue for eliminating one or the other indicator from the QRIS, as they may not be providing unique information (although some QRISs include certain quality elements to ensure that they are paid attention to for other policy related reasons, even if their psychometric properties are not ideal). To date, the Maine and Indiana validation research has not included a comparison of measures internal to the QRIS rating systems, but this is recommended in future research as the systems mature and stabilize.

Measurement error presents another potential challenge in assessing QRIS validity. Most QRISs assume that observational measures are relatively stable over time absent quality improvement efforts. This assumption is consistent with empirical evidence for at least one widely used instrument, the ERS (Clifford, 2005).

A related measurement issue concerns inter-rater reliability. In the twenty systems reviewed by Tout et al. (2010), nearly all QRISs require 80–85% agreement with a master coder (either exact agreement or agreement within one scale point) on ERS; this degree of reliability does not eliminate errors in ERS measurement (Bryant, 2010; Bryant, Burchinal, & Zaslow, 2011). For instance, two raters could be 100% reliable under a standard of 85% agreement within one scale point, but one might give a classroom a score of 3.5 and the other a score of 4.5, a difference that is large enough to affect an overall program rating (Karoly, Zellman, & Perlman, 2013). Based on the range and degree of variability in ERS quality scores at each rated level in both Maine and Indiana QRISs (see Tables 3 and 4), we recommended that program managers strive to increase the reliability of the rating process by clearly defining quality indicators and rating procedures, and conducting regular reliability checks.

Approach 3: assess the outputs of the rating process

A third validation approach focuses on assessing the outputs of the rating system: the scores and levels assigned to providers who undergo a rating, and the distributions of those scores within and across different types of providers. Studies conducted under this approach examine the degree to which the quality levels in the QRISs are meaningfully distinct from each other. The results of these studies may provide data that suggest that measures, cut scores, or rules for combining measures need to be changed in order to distinguish the rated quality levels effectively. Because these studies can result in proposals for significant changes to the standards for QRIS levels, it is helpful for these studies to occur prior to studies that examine associations between quality levels and children’s development.

Output studies may focus on individual indicator scores, such as how providers score on an environmental rating, as well as on the overall quality level that is the final output of the rating process. These studies may also utilize a measure of quality not included in the QRIS rating process to make an evaluation of concurrent validity, by examining whether assessments on both measures co-vary in predictable ways. The following section provides examples of the two states’ examinations of the distribution of quality ratings and rating-level advancement patterns for each program type enrolled in the state QRIS.

Examining initial QRIS rating distributions and cut points

While evaluators in Indiana and Maine did not conduct a detailed examination of the weighting or internal consistency of specific quality indicators, they did analyses to reveal the distribution of quality levels. After three years of system implementation, both Indiana and Maine QRIS child care providers were predominately rated at Level 1 or Level 2 (see Fig. 1). It is important to note that in Indiana, all providers enter the system at Level 1, and in Maine, providers can enter the system at any level based on their program rating, and then may advance at will from that level. A recent in-depth study of five state quality rating and improvement systems that were fully implemented found a similar pattern, with four of the five states reporting 40–76% of all programs enrolled in the lower tiers of the system (Mathematica Policy Research, 2011).

In Indiana, licensed child care centers were evenly distributed across the four QRIS levels approximately two years after the program inception. However licensed family child care homes were most frequently found at Level 1, with steeply declining numbers at the other three levels. This higher proportion of Indiana licensed centers rated at Level 3 or Level 4 may have been due to a greater historical emphasis in child care centers than in homes on regulation and attaining national accreditation, greater organizational capacity to complete the requirements of advancement in QRIS, or possibly that QRIS standards more closely reflect center quality than family child care home quality. Unlicensed registered child care facilities, a unique type of child care center in Indiana that is not licensed due to religious affiliation, participated at a much lower rate, and none had yet attained Level 4, reflecting significant challenges facing these unlicensed centers in meeting the Level 1 standards needed to enter the QRIS. These data patterns in Indiana
supported the validity of the QRIS rating system in that they showed variation in quality ratings across participating providers, they reflected the increasing effort necessary to meet quality standards at higher levels, and they were interpretable within the state’s child care context.

Another gauge of overall quality rating system utility is the amount of program advancement to higher rated levels. It is reasonable to expect, if the QRIS is viable, that at least some providers will advance in quality level. In the Indiana evaluation, 19% of the licensed centers, 24% of the licensed homes, and 27% of the unlicensed ministry centers had advanced at least one QRIS quality level in a 6-month period between assessments, during which mentoring was provided by local training providers. This advancement pattern, if maintained over time, suggests that even though attaining the highest levels may be challenging, quality improvement is feasible.

For Maine, as Fig. 1 illustrates, center-based programs and family child care type programs are most frequently found at Step or Level One. A disproportionately small number of family child care programs have attained Step Four, the highest quality level, and a disproportionately large number of child care centers and Head Start programs are enrolled at Step Four. This pattern of fewer child care homes enrolled at higher Step levels has existed throughout QRIS implementation in Maine. Maine family child care home providers argued that some of the program standards were “not a good fit,” despite designers’ beliefs that standards were well-matched to setting type. The large number of center-based and Head Start programs at the higher Step levels was expected, given that QRIS quality standards closely align with accreditation standards, and center-based programs are more likely to be nationally accredited than family child care homes.

An assumption of the designers of the QRIS in Maine was that programs engaged with QRIS will improve their tier levels consistently over time (Lahti et al., 2011). Approximately 80% of all programs (n = 1118) in the QRIS observed during the study period 2008 through 2011 did not experience a move up from one Step Level to the next. Results indicated that 95 of the 103 events or changes in Step Level from level one to two occurred during the first 23 months of enrollment. Moving from a Step One to Two, center-based care programs had a hazard probability of just 0.02 while family child care homes stayed virtually flat during this early period of enrollment in the QRIS. For movement from Step Two to Three, neither program type (p = 290) nor regional location (p = 195) appear to be significant in explaining Step level movement. For movement from Step Three to Four, the highest tiers in Maine’s QRIS, the analysis indicated that only type of program is a significant covariate explaining advancement. Family child care homes appeared to have a significantly lower probability of advancing a Step Level at this highest quality tier, compared with center-based and Head Start programs. These types of analyses of program movement in the system are relevant to the validation of a QRIS as they illustrate whether or not the way the system as designed is meeting its goals of supporting program advancement, leading to statewide improvement of program quality.

Studies may also be conducted to examine the degree to which given measures relate to other measures that purportedly assess the same concept. Here, strong correlation is desired, as they suggest that measures are measuring the concepts that they purport to measure in ways that are consistent with other measures of the same concepts.

**Indiana: assessing the output of the rating process**

The Indiana evaluation research included one validation test of state committee-generated quality standards, indicators, and levels by comparing the outputs of the QRIS rating system with independently gathered assessments of quality using validated quality measures, the environmental rating scales (ERS, Harms et al., 2005, 2006, 2007) and the Caregiver Interaction Scale (CIS; Arnett, 1989). The results, originally published by Elicker et al. (2011) and shown in Table 3, indicate that ERS scores co-varied as expected with QRIS level ratings, with a significant mean difference in global scores of 1.1 scale points between Level 1 and Level 4. Table 5 shows that caregiver interaction as observed using the CIS was less related to the rated QRIS quality levels. The overall correlation between the 4-level QRIS ratings and global ERS quality scores was moderate (r = .42, p < .01). The correlation between CIS adult–child positive interaction scores and QRIS level was more modest, but positive and significant (r = .24, p < .01).

Taken together, and looking across all types of providers, these results suggest that the QRIS ratings distinguish levels of quality in somewhat similar ways as two time-tested, validated measures of child care quality. However, mean quality levels at Level 4 were mostly found to be below the “good” rating threshold, suggesting the need to strengthen standards and/or rating procedures at the highest QRIS levels. In addition, finer analysis of the data suggested specific recommendations about quality standards and rating procedures that might be improved for each type of child care. Summaries of individual ERS item means for Level 3– and 4-rated providers led to the identification of a number of ERS items with scores below 4. Program planners are currently improving standards and QRIS rating procedures in light of these findings (Elicker et al., 2013).

In Indiana, patterns of association between QRIS ratings and ERS ratings were not the same for all types of child care. While the

---

Table 3

Indiana QRIS: mean global quality ERS scores\(^a\) as a function of program type and rated quality level.

<table>
<thead>
<tr>
<th></th>
<th>Level one (n = 84)</th>
<th>Level two (n = 90)</th>
<th>Level three (n = 74)</th>
<th>Level four (n = 66)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All providers (N = 314)</td>
<td>3.2 (.87)</td>
<td>3.7 (.76)</td>
<td>3.8 (.73)</td>
<td>4.3 (.80)</td>
</tr>
<tr>
<td>Family child care homes (n = 167)</td>
<td>2.9 (.64)</td>
<td>3.4 (.75)</td>
<td>3.6 (.67)</td>
<td>4.0 (.89)</td>
</tr>
<tr>
<td>Licensed child care centers (n = 133)</td>
<td>4.0 (.77)</td>
<td>4.0 (.68)</td>
<td>4.3 (.66)</td>
<td>4.5 (.67)</td>
</tr>
<tr>
<td>Unlicensed registered child care ministries (n = 14)</td>
<td>3.2 (.95)</td>
<td>4.1 (.45)</td>
<td>4.0 (.18)</td>
<td>NA</td>
</tr>
</tbody>
</table>

\(^a\) Possible range = 1–7.

Table 4

Maine QRIS: mean global quality ERS scores\(^a\) as a function of program type and rated quality level.

<table>
<thead>
<tr>
<th></th>
<th>Step one (n = 82)</th>
<th>Step two (n = 99)</th>
<th>Step three (n = 79)</th>
<th>Step four (n = 82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All providers (N = 342)</td>
<td>3.7 (.77)</td>
<td>3.9 (.84)</td>
<td>4.0 (.80)</td>
<td>4.3 (.79)</td>
</tr>
<tr>
<td>Family child care homes (n = 129)</td>
<td>3.3 (.67)</td>
<td>3.5 (.80)</td>
<td>3.8 (.91)</td>
<td>4.2 (.83)</td>
</tr>
<tr>
<td>Licensed child care centers (n = 165)</td>
<td>3.9 (.72)</td>
<td>4.1 (.80)</td>
<td>4.2 (.68)</td>
<td>4.4 (.84)</td>
</tr>
<tr>
<td>Head start centers (n = 48)</td>
<td>NA</td>
<td>NA</td>
<td>4.1 (.75)</td>
<td>4.5 (.71)</td>
</tr>
</tbody>
</table>

\(^a\) Possible range = 1–7.
global ratings were significantly correlated in both licensed centers and licensed family child care homes, the strength of association was stronger for homes, meaning QRIS level ratings in homes more clearly distinguished levels of ERS-related quality, meaning at each QRIS-rated level, the ERS quality differences were generally greater than they were for centers. Second, the overall ERS quality levels for center-based preschool classrooms (using ECECRS-R; M = 4.6 at Level 4) were somewhat higher than for center-based infant classrooms (ITERS-R; M = 4.4 at Level 4) and family child care homes (FCCERS-R; M = 4.0 at Level 4). While the equivalence of quality scores across these three ERS scales is not supported by research evidence, the results taken together suggest the need to strengthen quality standards and assessment procedures for all types of care, so that child care providers at the highest rated levels are providing care that is at or above threshold levels recommended to impact children’s developmental outcomes (Zaslows, Martinez-Beck, Tout, & Halle, 2011).

Maine: assessing the output of the rating process

As in Indiana, differences in program quality were measured using Environmental Rating Scales (ERS) mean scores at the classroom level, and these scores were not part of the QRIS standards or ratings. The results presented here are from factorial ANOVAs to examine the effects of Step Level, ERS scale type and child care program type on the dependent variable ERS mean score. Table 4 provides the adjusted mean scores for all 307 classrooms and by each program type by Step Level. Table 4 is extracted from the full report on Maine’s QRIS (see Lahti et al., 2011).

The results show an overall significant difference between Step Level and ERS mean score at the classroom/setting level (F = 5.02; df = 3, 307; p = .002). Results of post hoc Bonferroni tests showed a significant difference between Step One and Step Four programs (p = .001) and between Step Two and Step Four programs (p = .001). The total variance of the mean ERS score explained by Step Level was only 5%, indicating weak relationships between the variables. Comparisons of the program type mean ERS quality scores indicated a difference only between the family child care homes and the center-based scores (p < .001). The family child care home mean scores were lower at each Step Level than the center-based setting scores with the exception of scores at Steps Three and Four. There did not appear to be any significant differences at Step Three or Four between the center-based and Head Start type settings (p = .97). The results provide some evidence for differences in rated quality, with higher ERS means for higher tier or step programs, most distinctly for family child care homes. Overall these mean scores suggest the need for considerable efforts at quality improvement, considering that the majority of settings are scoring below the “5” or “good” level on the ERS measures.

Maine: parent level data on QRIS program quality

In the Maine validation study, parents in programs selected for observation were asked to complete an anonymous survey that focused on services received by the parent and the parent’s perceptions of the quality of the program. The belief was that parents served by higher Step level programs should be receiving more supports and services and therefore may rate the program higher in level of quality. The response rate over the three-year study period was approximately 26% (N = 1478). These results are extracted from the full report on Maine’s QRIS (see Lahti et al., 2011).

Parental perception of program quality was measured by the 15-item Emlen scale, see (Emlen, Koren, & Schulze, 2000), and was found not to be correlated to Step Level rating (Pearson’s r = .010, p = .68). In terms of services parents should have received according to program standards, a majority of parents reported not receiving: information about other government services for their child; opportunities for parent engagement with the program; daily communication from the program about their child; and being provided an up to date written parent hand-book from their provider. There did not appear to be any difference in step level in terms of parents not consistently receiving these types of services as required by the program quality standards according to parent reports. The use of these data by QRIS administrators was primarily for monitoring purposes focused on services and or supports parents should have received based on requirements in the QRIS standards. While parents were asked about perceptions of program quality, due to a strong desire to reach out to parents as a key stakeholder in the QRIS, that information was not relied upon for program planning or program improvement.

Approach 4: relate ratings to children’s development

The fourth approach to validation focuses on children’s development. In many respects, this is the final step in validating a QRIS, and one that arguably should be delayed until the questions raised in the earlier approaches are addressed and changes made to the system as necessary. It may even be possible that new data will emerge that makes the costly and difficult effort involved in assessing child outcomes unnecessary. For example, if studies begin to show consistently that certain inputs, e.g., ratings-based coaching lead to substantial improvements in indicators such as instructional support, and if instructional support or other indicators is found to consistently promote improved child outcomes, it may be possible to argue that the inclusion of those inputs and measures of those outputs may suffice.

The logic models that underlie QRISs typically assert that higher quality care will be associated with improved child outcomes. Therefore, one important piece of validation evidence concerns whether children make greater developmental gains in programs with higher program-level QRIS ratings than in programs with lower ratings. While a definitive evaluation of QRIS impact on child outcomes would consist of an experimental study with random assignment of providers and children to QRIS levels, ethical and practical considerations often make experiments impractical, at least on a state-level scale. Instead, current studies evaluating QRIS validity in terms of child outcomes using this approach do not attempt to evaluate causal linkages. Instead, they examine whether the QRIS ratings and the quality components that comprise the ratings are associated in expected ways to measures of children’s development. Showing significant associations between QRIS-rated quality would be a first step, a necessary but not sufficient result to demonstrate causal inferences about how QRIS quality influences children’s outcomes.
To date, few QRIS validation studies have incorporated children’s outcomes. Maine did not include this approach to validation. As Elicker and Thornburg (2011) note, results from such studies are mixed, at least in part because of the challenges of conducting them. A primary challenge is the inability to control for all the factors that may confound the quality-outcome correlations for children whose families have selected programs in a non-random way. Additional challenges include the difficulty of recruiting of programs and children across all quality levels; lack of information about the amount of care children received in each setting (dosage); lack of appropriate outcome measures for children of diverse ages, abilities, cultures and linguistic backgrounds; and, lack of variation in the quality of participating QRIS programs. As noted above in the discussion of Approach 3, measurement error remains a problem.

**Indiana: examination of ratings associated with children’s outcomes**

To examine validity-related questions about children’s development in the context of the Indiana QRIS, the evaluators assessed the developmental status of 557 children (249 infants/toddlers; 308 preschoolers) who were in the care of QRIS providers. Two children per classroom or home were randomly selected in approximately equal numbers at all four QRIS levels. Data from parent interviews describing annual family income and parents’ education levels and participation in the CCDF voucher program were used as control covariates in the analyses. The basic validity question explored was: are children in higher-rated QRIS care functioning at higher levels, socially and cognitively, than children in lower rated care?

It is important to point out that this study of quality and child outcome associations was cross-sectional, with all data collected at one point in time. As mentioned earlier, exploring these correlational relationships does not substitute for longitudinal or experimental designs that can better evaluate the causal impact of the QRIS on child outcomes. However in the implementation phase of QRIS, it is useful to explore the developmental status of participating children, how they are distributed in the child care system, and whether associations between quality measures and measures of children’s functioning are occurring in the expected direction (Elicker & Thornburg, 2011).

Bivariate correlations and multiple regression models were used to explore the associations between children’s development and the three measures of child care quality: QRIS ratings (4 levels); ERS global quality scores; and a CIS composite rating of positive adult–child interactions. All regression models included parent education, household income, and type of child care. No significant correlations were found between the four-level QRIS quality ratings and either infant/toddler or preschooler developmental status. Some of the researcher–observed quality measures were mildly but significantly correlated with child development measures. For preschoolers, CIS positive interactions were correlated with social competence \( r = .17 \), receptive language ability \( r = .17, p < .01 \). For infants and toddlers, ERS global quality scores were associated with social competence \( r = .15, p < .01 \), and total CIS positive interactions were associated with cognitive/language competence \( r = .17, p < .01 \). These significant correlations were entered as predictors in regressions of child outcomes on the quality variables, controlling for the family SES variables (parent education level and household income) that were also significantly correlated with the child outcome variables. As a result, for preschoolers, CIS positive adult–child interactions significantly predicted children’s receptive language ability, after controlling for family SES \( b = .12, p < .05 \). For infants and toddlers, CIS positive adult–child interactions significantly predicted children’s cognitive/language competence, after controlling for family SES \( b = .14, p < .05 \).

Family income was also a significant predictor, \( b = .23, p = .009 \).

Therefore while QRIS rated levels were not significantly associated with any child development measures for either infants/toddlers or for preschoolers, ERS and CIS quality measures were moderately associated with aspects of children’s development. Specifically, after controlling for family SES, it was the positive quality of interaction between adults and children that was associated with language and cognitive functioning, for both preschoolers and infants and toddlers.

Therefore it appeared that the specific aspects of child care quality assessed by the ERS and CIS measures are more likely to be associated with children’s development than are the composite of quality indicators represented by the 4-level QRIS ratings. This was true even though the QRIS ratings and the ERS and CIS were significantly correlated with each other. As a result, in Indiana, further refinement of QRIS standards and procedures is taking account of these findings, especially by identifying ways to strengthen QRIS standards and ratings to include the quality of adult–child interactions.

**Discussion**

**Limitation to validation study designs**

Both of these state studies provide results that describe linear associations among variables. The study designs are limited due to the fact that the investigators have no control over how the QRIS systems are implemented which affects enrollment and therefore sample sizes and selection of measurement strategies were also not in the sole control of the investigator. It will be interesting as additional studies are done and where non-linear associations are found to determine the impact this has on outcomes. These field studies were conducted with all the limitations associated with working in a developing system with multiple stakeholders. While the design presents a limitation in terms of arguing for causality and application of more sophisticated analytic approaches, it should be noted that the state agency program managers and other stakeholders in both states found the information generated from these studies of high value in terms of system planning, program improvements, and resource allocation. Depending upon context and resources, limitations to these two study designs can be remedied in future studies by such design choices as having programs that are on a waiting list be compared to programs already participating in their state’s QRIS.

**Validation of QRIS is a process that needs attention over time, using more than one approach**

The examples from Indiana and Maine illustrate how these validation approaches can work in practice, with tangible benefits for system improvement. These validation activities are specific to the design and implementation of each state QRIS. We believe it is important to stress to QRIS policy leaders that each of the four validation approaches needs to be used appropriately, considering the developmental stage of the QRIS and the unique features of the setting and QRIS. For example, states with QRIS in development can use the four approaches as a framework for planning how to validate their system. Developmentally then, an initial focus of QRIS design would be to validate the key concepts used in the QRIS design. The four approaches highlighted in this paper need to be considered as part of an ongoing process, not a one-time event. As states progress in their implementation of QRISs, more descriptive research is necessary to understand better how these validation approaches work in other settings, for example with point-based QRISs (rather than “building block” QRISs, like those in Indiana and Maine). Use of these approaches enables cross system comparisons,
which will allow for the identification of common threats to validity and useful strategies to enhance the validity of a state QRIS.

Validation and early care and education system constraints

A QRIS is not merely a program-level quality-improvement intervention, it is a policy lever for strengthening a state's overall early care and education system that reaches beyond child care (Schaack, Tarrant, Boller, & Tout, 2012). The two state validation efforts highlighted in this paper reflect the challenges and constraints common to other state experiences with validation activities (Lahti, Sabol, Starr, Langill, & Tout, 2013). On-site observations of global program quality and establishing and maintaining inter-rater reliability for QRIS raters is a time-consuming and costly endeavor. Keen interest in school readiness may pressure program administrators to collect child outcome level data before a QRIS is well established. Current research on the measures that are in the widest use to predict child outcomes appears to do so consistently, especially for children at-risk, but with modest levels of association with program quality as measured by the ERS's (for example, Burchinal et al., 2011). We recommend to policy makers to always take into account that any validation study is occurring within a dynamic system. System-level constraints such as varying resources available to programs, different type and design of programs, and challenges to measuring quality and reliably collecting information about program quality all influence the design and implementation of state-level QRIS validation studies.

Validation research is critical for performance measurement and improvement for a state QRIS

The 2012 Child Care Development Fund (CCDF) Plan preprint for fiscal years 2014–2015 includes a much larger focus on QRISs (U.S. Department of Health & Human Services, Administration for Children and Families, 2011). In this document, a QRIS is defined as a “...systematic framework for evaluating, improving, and communicating the level of quality in early childhood programs.” States are expected to provide a self-assessment based on current program quality initiatives from a set of questions that are also organized according to a “QRIS framework.” Validation of program standards or assessment tools is mentioned specifically in relation to information states must provide about data and performance measures on program quality.

The information generated from QRIS validation activities can be used to inform efforts for continuous quality improvement. For example, both Indiana and Maine found that, for at least some types of providers, enrollment patterns in the QRIS and lack of movement by programs once they are in the QRIS, is resulting in a large proportion of providers at the lower-rated quality levels of the quality tiers. System-level, quality-improvement responses to this information could be to re-assess the design of the system in terms of the ability of programs to meet standards at each tier, or to focus training/technical assistance on specific quality standards that are most challenging for providers to meet. At the same time, care should be taken to ensure that standards reflect current knowledge about the specific indicators and levels of quality most likely to produce the desired child outcomes. Findings from validation studies can be part of the information that state child care administrators use to assess the overall performance of the state early care and education system. The performance data could then be used by program administrators in making decisions about monitoring programs in a differential manner by visiting those programs more often who are having difficulty meeting QRIS standards (Fiene, 2013). We recommend that one focus of future research be learning more about whether and how information from the results of validation studies are used to improve system and program level performance in QRIS.

It will be interesting to determine as more validation studies are completed to analyze the differences between levels and how often lower quality is present in the top level which is the case with measuring compliance with licensing standards (Fiene & Nixon, 1985). Key areas to look at will be the movement of programs from one level to another, how long this takes, and are the increments equal or not in terms of quality improvement.

Another area to be explored which may have an impact on overall QRIS implementation are the fiscal constraints that many states are experiencing due to the recent recession and lower levels of federal funding. It would be interesting to note differences amongst states with large investments in quality improvement initiatives and those states with smaller investments.

It is important to remember that the QRIS is a policy lever and the validation of child care quality standards in a QRIS is a new phenomenon in early care and education policy-making. These four recommended approaches to QRIS validation, illustrated by validation research in these two states, even with their limitations, did provide policy makers and program administrators with information that guided efforts at system quality improvement. The use of these approaches in other studies will create a common nomenclature for better understanding threats to validity in a QRIS and ultimately increase our understanding of how best to design a QRIS that meets the needs of the parents, providers and children it serves.

References


Please cite this article in press as: Lahti, M., et al. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. Early Childhood Research Quarterly (2014), http://dx.doi.org/10.1016/j.ecresq.2014.04.005


Elicker, J., Langlil, C. C., Ruprecht, K., & Kwon, K. (2007). *Paths to QUALITY*: A child care quality rating system for Indiana. What is its scientific basis? Lafayette, IN: Purdue University, Center for Families.


Mastering course content and learner satisfaction in early childhood education: A comparison of regular classroom instruction with three variations of internet delivery

James Johnsona,*, Richard Fienea, Jane Keata, Harriet Darlinga, Donald Prattb, Joyce Iutovichb

aThe Pennsylvania State University, Pennsylvania, USA
bKeystone University Research Corporation, USA

Abstract

In the spring semester 2000, a Penn State course, ECE 479 (The Young Child's Play as Educative Process), was taught by the same instructor in four delivery formats. One group consisted of a regular classroom, held on campus. A second group, also on campus, was taught in a computer lab via the Internet; and there were opportunities for interaction with peers and the instructor. A third group took the course on the Internet as part of a local distance education group; hence, there were some limited opportunities for face-to-face interaction with peers and the instructor. The fourth group took the course on the Internet, as part of a statewide distance education group, where there were no opportunities for face-to-face interaction. Twenty students who enrolled in the course (5 per group) completed questionnaires and phone interviews. Information was gathered on professional backgrounds, computer experience, and initial level of content knowledge on the topic of the ECE Internet course. Sixteen students who completed the course were interviewed again to evaluate satisfaction with the course and to estimate learning outcomes. Across the four conditions general satisfaction was expressed with the content, activities, and course requirements and with the teacher. However, students in the three computer groups expressed dissatisfaction over technical problems (all four who did not complete the course came from these computer groups). Significant gains in content knowledge occurred for the classroom group, while the learning in the three Internet-based instruction groups did not show the same gains. Concern was expressed related to the lack of face-to-face interaction, making the learning environment less desirable. Although Internet technology provides a great deal of promise, these results suggest that improvements are needed to make this delivery modality more effective for in-service distance learning. © 2001 Elsevier Science Inc. All rights reserved.

1. Introduction

"I think the thing that I liked least and I know it's inherent in an Internet course is lack of human involvement. Lack of human contact. Not getting to talk with the instructor at all, face to face. That was very tough for me. The other thing was the lack of the immediate feedback. If you had a question you had to type it in and wait for the instructor to log on and read it and then get back to you."

*This research was funded by Keystone University Research Corporation and the Penn State Consortium for Children, Youth, and Families. The authors gratefully acknowledge the role played by ECE doctoral students Hey-Jun Ahn and Juli Grabinak in helping develop the on-line course and World Campus specialist Cathy Holsing for her technical support and encouragement.

Corresponding author. Tel.: +1-802-814-865 ext.2230; fax: +1-814-865-7602
E-mail address: jej4@psu.edu (J. Johnson).

0169-6288/01/$ see front matter © 2001 Elsevier Science Inc. All rights reserved.
PII: S0169-6288(01)00129-5
"My frustration was that I don't like to fall behind in anything. I felt like I was falling behind because I wasn't able to access these readings, wasn't able to do some of the things on the computer that I should have been able to do. I was just getting frustrated and feeling like this isn't worth it for me. And that's why I dropped the course."

"The lack of being able to talk to anybody. The fact that you were doing all of this reading on your own. That was hard. I was used to sitting in classes and having someone lecture to me and then having students to interact with. It was difficult to get used to just reading it and absorbing it by yourself."

"Just simple things like just being able to print the page on my screen out. I didn't, I had no clue how to transfer stuff onto the hard drive and to be able to print something. I'm used to pressing the print button and stuff popping out. I didn't know it was that much involved."

The Internet is a new and ever expanding tool for learning (Benson & Meyers, 2000; Lan, 1999; Ovston, 1998), including distance education (Williams, 1996). Whether or not the Internet can be an effective tool for training staff within the human services, particularly childcare staff, needs to be explored. Our project evaluated the effectiveness of the Internet training in terms of learning outcomes, its implementation (specifically, the technological aspects), and the student's level of satisfaction with the course. The impetus for our project came from two sources.

First, Pennsylvania delivers childcare training to all licensed and registered childcare providers in the state and is interested in making this system more cost effective and efficient. Discussions related to utilization of the latest technologies, such as the Internet, are being considered. Secondly, a new initiative, CyberStart, will link all licensed childcare centers in Pennsylvania to the Internet. While this initiative is specifically designed to offer Internet access and educational programming for children, it will also make this technology available to childcare staff. Hence, there is a need to evaluate the feasibility and effectiveness of Internet-based distance education as it becomes more available to childcare centers.

2. Research design

A quasi-experimental design was employed, which consisted of four groups of five students who enrolled in the early childhood education (ECE) course ECE 479 The Young Child’s Play as Educative Process, which focused on play, communication, and curriculum. The first group experienced the traditional lecture/discussion course format. A second group took the course on the Internet, but within the context of a computer lab (located at Penn State York) where they had the opportunity to interact with their peers and the instructor. A third group took the course on the Internet as part of a local distance education (DE) group; this group also had some face-to-face interaction with their peers and knew the instructor. A fourth group took the course on the Internet, but as part of a statewide DE group. This group had no face-to-face interaction with their peers or the instructor.

This research design enabled us to examine the available technology to determine any hardware or software constraints, as well as the efficiency of the technological support services, by comparing groups that took the course via the Internet in different environmental circumstances (i.e., the on-campus computer lab vs. a home-computer set up). The research design also permitted us to evaluate the importance of the human element as a component of the effectiveness of this training modality since participants were in controlled settings with varying possibilities for face-to-face interaction.

Evaluating whether Internet training delivered to childcare staff is an effective training modality has special significance at this time given the paucity of empirical evidence to support or refute such claims. Research is needed to help the Pennsylvania Childcare and Early Childhood Development (PACC/ ECD) Training System, and comparable systems in other states, make more informed decisions about the use of Internet-based distance education for teachers of young children.

Given the exploratory nature of this study, a qualitative approach with a small sample was employed to generate data from questionnaires and interviews, as well as from course assignments completed by students. The questionnaire included items that tapped demographic characteristics (age, sex, and prior education), current position and experience within childcare, and experience with computers. Phone interviews, administered before and after the course, and lasting in duration from 30 to 45 min each, assessed students' knowledge about play and perceptions about the course. In addition, select course assignments were independently graded to assess knowledge.

The course aimed to increase the students' knowledge about play and its practical application. Two measures of the learning outcomes were used: (1) interview responses to questions about play given before and after the course and (2) grades on selected course assignments. The latter were independently evaluated by faculty other than the course instructor. Both measures were scored without knowledge of group membership.
From the phone interviews, students' answers to four questions about play were evaluated. The questions were: (1) What is play? (2) What is the value of play? (3) What is positive play? and (4) How can adults have a positive influence on play? Content analysis revealed that numerous ideas were elicited by these four questions both before and after the course. An empirically based coding system was developed and employed to score students' answers. Appendix A provides the categories and subcategories used in scoring.

3. Course development

The course The Young Child’s Play As Educative Process (ECE 479), offered by the Department of Curriculum and Instruction in the College of Education at The Pennsylvania State University, was selected for this experimental study of Internet-based instruction. This particular course was chosen because it was viewed as well-developed and because it included both theoretical as well as practical content relating to ECE curriculum and instruction. Developing the course for Internet-based instruction required connecting with the Penn State World Campus (online learning). A contractual relationship was formed involving Keystone University Research Corporation (KURC), the contractor for the PA CC/ECD Training System, the College of Education, and the World Campus. Technicians from the World Campus, in collaboration with ECE faculty and doctoral students, worked to develop the course in time for offering it in the Spring of 2000.

3.1. Stage one – course structure

One of the first tasks was to establish the course syllabus, which included the course objectives and requirements, along with a schedule and sequence of learning activities. Objectives included acquiring a knowledge base of theory, research and practice concerning play and ECE, improving play observation and documentation skills, analyzing playthings for use in ECE settings, planning suitable indoor and outdoor environments for play, understanding and making sound judgments about the use of play facilitation strategies, and learning how and why to become an advocate for play in education. Requirements included an observation project, designing play environments, writing letters explaining play-based teaching to parents and to a 'blue ribbon committee' of educational professionals, as well as doing an implementation project and keeping a journal. The Instructor’s Manual to Accompany Johnson/Christie/ Yavkey Play and Early Childhood Development, Second Edition (Johnson, 1999) provided the guidelines for the overall course organization and sequencing, which followed the chapters of the text, with the content going from theory and research to policy and practice.

3.2. Stage two – course content

The course was organized into four modules with a number of online lessons or sessions in each module. There were a total of 26 sessions. Sessions designed to last 50 min; students had a reading assignment for each session and a self-administered objective-item exam, which produced computer-generated feedback for self-evaluation. Also, open-ended discussion questions were assigned for chat room or bulletin board discussion. The objective items and the open-ended discussion questions came from the instructor’s manual. Activities and problems were embedded in sessions and were sequenced and used in the same way across the four groups of students.

3.3. Stage three – course programming

The final formatting and programming of the course for Internet instruction occurred during the Fall of 1999. The course program included some special features to make the on-line learning experience more interesting. For example, an animated pop-in character (a cartoon owl) appeared on screen at various selected points throughout the sessions to ask questions as a real classmate might. The World Campus technical staff, with the assistance of the course instructor (third author), also prepared a home page for the course, which included the course schedule and contact information.

4. Participants

All 20 students were female, ranging in age from 23 to 60 with mean of 39 years. Sixteen students had a bachelor's degree and one a master's and two had a high school diploma or GED certificate. The most common college major was elementary education (N = 7). Only three students majored in ECE and had taken any previous courses on play. For the entire group, inservice training hours on this subject ranged from zero (N = 5) to 45 hours (N = 1), with a mean of 10.5 hours.

Fifteen students held the position of center director and three were assistant directors. One student had the title of childcare coordinator, and another was a personnel training coordinator and an assistant group supervisor. The length of time that the students worked at their current childcare center ranged from
5. Findings and discussion

5.1. Prior computer experience

Students reported prior experience using personal computers ranged from no experience to 15 years. The mean value was 4.1 years. Seventeen students had a personal computer at home and were asked to indicate which specific activities they had used. Those uses included: word processing (15 responses); recreational software (12 responses); Internet (12 responses); and spreadsheets and/or databases (4 responses).

A questionnaire item asked who usually provided technical support for their home PC (e.g., installing new software or hardware, answering software questions, and fixing problems). No more than five students usually relied on themselves for technical support for their home PCs. Of the four students who chose "other," three students usually relied on a friend, and one usually relied on another teacher in her childcare center.

5.2. Knowledge about play

For the following four play questions, each student’s answers given in the postcourse interview were compared with the precourse answers in order to evaluate whether there were response improvements, defined by an increase in positive responses and/or a decrease in negative responses (items 1 and 3), by an increase in the total number of distinct acceptable benefits of play given (item 2), or of distinct teacher’s roles (item 4) cited.

5.2.1. Question 1: what is the definition of play?

Responses were coded into the categories and subcategories and summed into two groups. Positive responses tallied the number of Attributes, Distinctions, and Realizations (A, D, R), while a second group of negative responses summed the total of overly-inclusive, mistakes, slogans, and vague replies (X, M, S, V). Other codes were considered "neutral" answers and were not used in this evaluation.

All classroom and computer lab students improved, as did the three statewide DE students who stayed in the course. Of the three remaining local DE group students, one failed to exhibit a positive response at both times of measurement; they gave five negative responses at precourse time of measurement and three negative responses at postcourse time of measurement.

Classroom students gave a total of 25 positive responses after the course compared to only 3 positive responses before the course. This averages five positive responses per student and exceeds the averages on this index for the other three groups (1.2, 1.6, and 2.3 for lab, local DE, and statewide DE, respectively). Classroom and lab students showed few negative responses before or after the course, while local DE students gave 8 negative responses and statewide DE students gave 13 negative responses before the course. However, in postcourse interviews the six DE students made only five negative replies.

5.2.2. Question 2: what are the benefits of play?

Total scores for number of distinct benefits were derived from the coding system employed in the study. All general and expressive answers given (G and E) were added with all functional answers to yield a total benefits (G&E + F) score. Benefits of play results are shown in Table 1.

As can be seen in Table 1, the classroom and lab students performed better than the local DE and statewide DE students on this item. They were able to identify significantly more functions of play, especially after the course was taken. Nine of ten students in the classroom/lab groups improved, but only 3 of 6 in the DE groups did. Note also slight decrease in scores by local DE group.

5.2.3. Question 3: what is good play?

Here positive responses were the number of Nominal and Process (N and P) ideas from each student, while a second group of negative responses summed the total of Failure (F) responses exhibited. Other codes were considered “neutral” and were not used here.

Only seven of 16 students improved. Evidently, the course did not impact very much how well students could answer this question. The best answer, the F answer, was given by only three students, all at postcourse time. Failure to distinguish good play from play (e.g., “all play is good”) happened six
times at pre-course time and eight times by the end of the course. Most students talked about types, functions and characteristics of play and failed to differentiate good play from play in an acceptable manner as determined by the coding and scoring system used in this study. This discouraging result may be because the scoring criteria were too harsh, or perhaps because interviewees did not understand the question. Also, it is possible that the course as taught, as well as the textbook as read, did not highlight the distinction well enough. In previous courses, this question has proven to be a "power question," distinguishing the "gold from the dross" in student performances.

5.2.4. Question 4: how can adults make child's play better?

Responses were analyzed with scores derived from the coding employed in the study. All general or attitudinal answers (G) per student were counted. Another group of answers covered the sum of all adult role responses (R). Total roles (G&R) were examined in the analysis. Table 2 shows mean scores across groups.

There were 11 students who improved, with all but one of them coming from the classroom or computer lab research groups. Only one student in a DE improved. Five students in DE groups actually performed more poorly in answer to this question after the course was over. In sum, the course influenced students' performance in identifying adult roles in children's play differentially depending on which research group they were in—classroom and lab students outperformed DE students.

5.3. Assessment of course requirements

There were three assignments that were graded for all the students: implementation activity (A), a parent letter (B), and a blue ribbon letter (C). As shown in Table 3, a comparison of the grades on these assignments across all four groups did not reveal that any one group consistently scored higher than the others. However, the traditional classroom group did score the highest on assignment B and had the second highest set of scores on assignments A and C.

For assignments A and B the differences in scores for the four groups were statistically significant (Assignment A, $F=5.574, p < .012$ and Assignment B, $F=4.628, p < .023$). No significant differences were found for Assignment C. The computer lab group scored the lowest on assignments A and B while the other three groups of students all scored about the same. When all the assignments are totaled for an overall score, the traditional classroom performed significantly better than the other three groups ($F=5.221, p < .015$).

Students were also asked a series of questions about their expectations regarding the course, the benefits from taking the course, and their experiences in taking the course. The majority of the responses were very positive about the course in general, but the students who had taken the course via the Internet experienced a number of computer problems in either getting online or with the software. Everyone experienced problems and this delayed the start of the course for them for several weeks. There were problems with signing on, with the chat room, with passwords not being accepted, and so forth. However, once these problems were worked out, the course started and proceeded fairly well.

In sum, respondents revealed general satisfaction with the content of the course, the course activities, and course requirements. Moreover, there was a strong appreciation and high evaluation for the teacher. On the other hand, almost everyone (one or two exceptions) in the three computer groups expressed dismay over serious and continuing technical difficulties throughout the entire semester. However, even with a good course content, instructional design, teacher, and technical delivery, there were several people who clearly indicated reservations about Internet learning because it lacked face-to-face interaction. These people indicated that even if there were no technical difficulties, they would miss the human contact and would prefer courses or training taken in a classroom where there was greater opportunity for interpersonal interaction and contact.
6. Conclusion

This evaluation provides insights into offering ECE courses over the Internet. Clearly, it seems that the success of this technology is dependent upon the persistence and knowledge of the student for learning to occur. Four students who did not complete the course were from the local (N = 2) and statewide (N = 2) Internet-based distance education groups; none were from the traditional classroom or computer lab settings. The students were administrators of childcare programs, fairly familiar with computers and knowledgeable of the course content. However, even with these pluses, the students still had considerable difficulty in accessing and doing the course online. Possibly the dropout rate would have been greater if the students were at a beginning stage of their career (Cohen, 2000; but see Schrum, 1992).

Students in the traditional classroom and in the computer lab groups, where there was more face-to-face interaction, scored the best on the interview play evaluations. Evidently, these two training modalities were more effective than were the local and statewide distance education training modalities where there was little or no face-to-face contact with other students or with the instructor.

On the three course requirements, the classroom group scored the highest on assignments summed together (9.47); but the computer lab group scored the lowest on the assignments (7.53), with the two distance education groups in the middle. This is inconsistent with the results from the analyses of the interview responses to the four questions about play. Here both the classroom group and the computer lab group gained the most. Although it is perhaps encouraging to see that the students in the two distance education groups scored higher on the course assignments compared to the computer lab group, their results were still lower than the students' scores from the traditional classroom group.

What have we learned from this study? For Internet instruction to be effective it seems that students must be technologically literate and knowledgeable about course content to some degree, and they must be persistent and highly motivated. Without these personal characteristics, the Internet course experience may not be a positive learning experience. The results of this study further suggest not utilizing Internet training across-the-board for childcare staff. It needs to be very targeted, beginning with directors of programs, who generally have the greatest experience and education and potential exposure to computer technology. More generic, across-the-board training for the beginning level practitioner does not make training sense at this point.

Internet technology provides a great deal of promise for reaching childcare staff with needed specialized in-service training in ECE. But first it is necessary that the technology gets the fine-tuning to ensure its effectiveness as a training modality for the majority of childcare providers. Also, what is best for on-line learning? Professional attention must be given to criteria for deciding the kinds of training content (e.g. learning facts like state regulations or understanding and applying concepts like play in ECE) in relation to its packaging (e.g., a full course like in this study, or modules, or single sessions analogous to workshops). Setting up realistic expectations for learners and designing the right means of instruction and assessment are both very important.

Finally, we recommend that priority be given to designing orientation modules to prepare potential users for Internet-based distance education.

References

William, A. (1996, June). Integrating courses with the Internet: Preparing the teacher as well as the learner. Paper presented at the summer conference meetings of the Association of Small Computer Users in Education, North Myrtle Beach, SC.
Appendix Table 1

**Question 1: What is play?**

A1 Attribute of play given. For example, play is process-oriented, marked by positive affect, often non-literal, active, intrinsically motivated, etc.

A2 Tangential attribute of play given. For example, play is something adults like to see in children; child and adult do play of own free will, etc.

E Example of play is given. For instance, play is like playing with dolls or blocks.

F Function of play is noted as a way to define it. A way to learn things. A way to release feelings.

C Context of play is noted. Play can be solitary or done in a social group.

X Overly inclusive or general statement about play is made, such as “It is a creative process.”

D Distinction is made, contrasting play with other similar behaviors such as imitation or exploration.

R Realization is indicated, such as play is multidimensional, complex, or hard to define.

T1 Type of specific play is given, such as constructive or dramatic play.

T2 Type of general play is noted, such as educational play, dark play, and recreational play.

M Mistake is made, a falsehood is uttered.

S Slogan is cited; such as play is the child’s work.

V Vague pronouncement is made, unclear or hard to decipher.

**Question 2: What is the value of play? What are its benefits?**

General (educational) values include:

G1 Teaches skills and abilities or is a learning experience

G2 School preparation

G3 Allows for the practice of skills

G4 Generates further learning and development; allows children to go as far as they can with what they are learning

G5 Avenue for creativity

G6 Creates a well-rounded child

G7 Allows child to explore and discover on one’s own

G8 Allows child to experiment

Expressive values include:

E1 Enjoyment, fun, makes happy, love of life

E2 Relaxation, release energy, tension

E3 Vent frustration

E4 Be self

E5 Express self

Functional values include:

F1 Cognitive: abstract thinking + imagination + learning content + creativity + learn on own terms-relevancy, meaningfulness + problem-solving + meta-cognition + memory + social cognition, empathy, perspective-taking + theory of mind + sense of self + sense of others + assimilation, integration, application of learning

F2 Affective: motivating, feel good about self + self-empowerment, sense of control + reduce anxiety, therapeutic, cathartic + self-confidence, sense of self-assuredness

F3 Social: learn to resolve conflicts + cooperation + group cooperation, team member + leadership skills + learn to share + learn to take turns + learn to help

F4 Physical: gross motor + fine motor + learn to challenge self physically + self-help skills

F5 Attentional: concentration + attention regulation + persistence

F6 Assimilation

F7 Language: communication skills + literacy + become good story-teller + vocabulary

F8 Academic: reading and writing + shapes for math + science

F9 Life skills, careers

**Question 3: What is good play? Positive play?**

N Nominal answer is given, like educational play is good play.

P Process of play is said to be important, “perking along”, playing up to capacity.

C Characteristic of play in general is given without really answering the question.

T Type of play in general is given without really answering the question.

F1 Function of play in general is given without really answering the question.

F2 Failure to provide sensible reply, such as, “all play is positive.”

(continued on next page)
Appendix Table 1 (continued)

**Question 4: How can adults make play better for children?**

There are many roles to perform in the ECE profession with respect to children's play. Some are general or attitudinal such as:

- **G1** Value play, allow it to happen, be aware of your biases and those of others
- **G2** Realize when attitude change about play occurs in self or others
- **G3** Show an interest in play as a matter of public or educational policy
- **G4** Lobby and advocate on behalf of children's right to play as an integral part of childhood

Other roles in the ECE profession with respect to children's play pertain to classroom or home or child development center situations where young children learn and are cared for. These include:

- **R1** Stage manager, set up over-all positive and attractive environments, provide materials, playthings, space, toys, storage areas, time for play, literacy props, allow children choice of areas and activities, organize settings, rotate toys
- **R2** Be a careful and systematic observer, evaluate play, change own behavior as a result of observations, document play process and products
- **R3** Provide preparatory experience, bring in guest visitors, field trips
- **R4** Play facilitator, scaffold, support, challenge children at play, do not be overbearing, ask questions, offer comments, suggestions, ideas for play, teach play
- **R5** Co-player, be a play leader, model play behavior
- **R6** Supervise play, make sure play is safe, monitor activities, mediate conflicts, enforce rules, help children negotiate, guide and discipline, be a referee

Other roles inherent in the ECE profession connected with play have to do not with the children but with other adults.

These include:

- **R7** Help adults be comfortable with play of children in educational settings, train teachers, educate staff about play in ECE
- **R8** Inform and involve parents in quality play with young children in formal and informal educational settings
Improving Child Care Quality Through an Infant Caregiver Mentoring Project

Richard Fiene
The Pennsylvania State University

ABSTRACT: An evaluation of a mentoring training program for infant caregivers is described. Fifty-two infant caregivers from 27 childcare center-based programs were involved in a four month long intervention in which they were paired with an experienced early childhood educator. The focus of the mentoring program was to improve the overall quality of the classroom environment, as well as making the caregivers more sensitive to the needs of the infants. The results clearly indicated that the mentoring program was very effective in improving the overall quality of the classroom, as well as making caregivers more sensitive to infants’ needs.

KEY WORDS: infant caregivers; childcare; mentoring; training.

Introduction

This paper describes a child care mentoring project designed to improve the quality of infant and toddler child care programs in south central Pennsylvania. The goal of the mentoring project was to improve the quality of the child care environment and specifically the quality of caregiver-child interactions. As most caregivers in Pennsylvania only receive workshop training, the goal of this project was to compare the mentoring approach to the more typical workshop training. Mentoring is being explored because of its targeted intensive one-on-one nature in delivering training to caregivers based upon needs assessments. The project was conducted during the later half of 2000 and the beginning of 2001. The results presented in this paper are part of the pre- and post-test data collection phase (summer 2000 and winter 2000–2001) of this mentoring project. The actual mentoring intervention occurred from September through December 2000.
Mentoring in childcare has been documented in the literature for the past 10–15 years (Breunig & Bellm, 1996; Fenichel, 1992). It has been demonstrated to be an effective mode of training/technical assistance (Breunig & Bellm, 1996). However, in the majority of studies conducted there are few, if any, demonstrations that utilize a randomized trial design (Breunig & Bellm, 1996). Many studies track the progress of the intervention group, some studies have comparison groups, but few, if any, have employed a randomized design. This research paper will describe the pre- and post-test data collected as part of a study that has employed a randomized design.

The majority of research (Clarke-Stewart, 1987; Goelman & Pence, 1987; Howes, 1987; Phillips, 1987; Kontos & Fiene, 1987; Galinsky, Howes, Kontos, & Shinn, 1994; Scarr, Eisenberg, & Deater-Deckard, 1994; Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997; Helburn, 1995; Fiene, 1995, 1996; Jorde-Bloom, 1988; Love, Schochet & Meckstroth, 1986) completed on early childhood quality has focused on pre-school programs, with infant toddler programs rarely as the central focus of the research. The research completed in infant toddler programs has clearly documented the mediocre level of care provided to children in these programs (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997). In the present study, we focus on the first three years of life. All the centers and the classrooms reported upon in this study serve children from birth to less than three years of age.

This report is organized as follows: a methodology section briefly describes the sample selected with basic demographic information on directors, caregivers and the programs. This is followed by a results section that provides pre- and post-test average scores for each of the assessment tools utilized in this study to measure quality, caregiver behaviors, knowledge, and organizational climate of programs. This section is followed with a discussion section and implications regarding this mentoring project.

Methods

Study Design

This study involved 52 caregivers from 27 sites in south central Pennsylvania. All programs were child care centers licensed by the Department of Public Welfare. Seven of the sites were accredited by the National Association for the Education of Young Children.

This study employed a randomized design in which a self-selected group of programs and caregivers were randomly assigned to two groups, either the mentoring group or the comparison non-mentoring
comparison/control group. Intervention model mentoring group received intensive mentoring from a seasoned early childhood professional (minimum of 5–7 years of experience in the early childhood field as both a director and teacher) from September to December 2000. The mentoring model consisted of a problem solving approach in which the mentor spent a good deal of time observing in the beginning weeks in order to develop a trusting relationship with the protégé. Once both the mentor and protégé felt comfortable then suggestions could be entertained by the mentor.

The comparison group did not receive the mentoring intervention and only had the regular workshop type variety training available to them. However, the comparison group did receive mentoring during the Spring 2001 from March to June 2001. What is of interest in this study is to determine how much the two groups have improved from the pre-test data collection because they were essentially equivalent at that point on all measures.

Programs were recruited by the Capital Area Early Childhood Training Institute, a broad based community focused training institute. Program directors were invited to attend a meeting describing the mentoring project. Of those attending, 95% agreed to participate in the project. Fifty two caregivers started the project, 14 caregivers dropped out of the project between pre- and post-test. There was an equal drop out rate from both the mentoring and the control groups.

Data from the four quality measures used for all the programs are presented in Table 1. The four measures of quality were the Infant Toddler Environment Rating Scale (ITERS), the Arnett Caregiver Observation Scale, the Knowledge of Infant Development (KIDI), and the Bloom Scales of Organization Climate.

The program directors’ average age is 31 with a range from 24–53.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>ITERS, Arnett, KIDI, Bloom Scale Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Programs (n = 38)</td>
<td>Pre-Test</td>
</tr>
<tr>
<td>ITERS</td>
<td>134</td>
</tr>
<tr>
<td>Arnett</td>
<td>30</td>
</tr>
<tr>
<td>KIDI</td>
<td>14</td>
</tr>
<tr>
<td>Bloom</td>
<td>78</td>
</tr>
</tbody>
</table>
years of age. They are predominantly Caucasian (81%). Eight percent have associate degrees, 78% have bachelor's degrees, and 14% have master's degrees. They had been employed as directors in their program for an average of 31 months with a range from 1 month to 120 months. Their average pay is between $20000–25000 per year. Sixty percent have health insurance and 45% have some form of dental or life insurance. Forty-five percent are in a retirement system.

The average age of caregivers in the programs was 36 with a range from 18–68. They are predominantly Caucasian (77%). Fifty-seven percent have high school diplomas, 16% have some college credits, 5% have CDA's, 16% have associate degrees, 5% have bachelor's degrees, and 2% have master's degrees. They have been employed as caregivers in their program for an average of 34 months with a range from 1 month to 153 months. They have worked in the early childhood field as caregivers for an average of 71 months with a range from 1 month to 312 months. Their average pay is between $10000–15000 per year. Fifty percent have health insurance and 33% have some form of dental or life insurance. Thirty-three percent are in a retirement system.

The average size of the centers is 98 children with 17 staff employed either full time or part time at the program. The average weekly fee for infant care is $137.00 per week and for toddler care is $124.00 per week. The majority of staff are employed at the centers for either less than 1 year or greater than 5 years.

**Results**

Both the mentoring and comparison groups were tested for equivalence at the beginning of the project in the pre-test data collection phase. There were no statistically significant differences on any of these measures at the pre-test. When the programs and caregivers were measured at the post-test, positive changes occurred although none were found to be statistically significant. In the aggregate, the programs that continued with the mentoring project showed improvements in the overall quality of care.

Tables 2 through 5 present the pre- and post-test data for the intervention and control groups.

These results indicate that the mentoring group showed increases on the program quality scales (ITERS and Arnett). This increase is especially noticeable on the ITERS. Further, there was a decrease in program quality with the control group, going from a score of 137 to 132. On the Arnett scale the mentoring group increased greater than the control group (11 point increase versus a 7 point increase).

Although the above results did not reach statistical significance,
Table 2
ITERS

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>134</td>
<td>141</td>
<td>+7</td>
<td>ns</td>
</tr>
<tr>
<td>Control Group</td>
<td>137</td>
<td>132</td>
<td>−5</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 3
Arnett

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>29</td>
<td>40</td>
<td>+11</td>
<td>ns</td>
</tr>
<tr>
<td>Control Group</td>
<td>33</td>
<td>40</td>
<td>+7</td>
<td>ns</td>
</tr>
</tbody>
</table>

Table 4
KIDI

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>14</td>
<td>14</td>
<td>-0-</td>
<td>ns</td>
</tr>
<tr>
<td>Control Group</td>
<td>14</td>
<td>15</td>
<td>+1</td>
<td>ns</td>
</tr>
</tbody>
</table>
Table 5
Bloom

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentoring Group</td>
<td>73</td>
<td>74</td>
<td>+1</td>
<td>ns</td>
</tr>
<tr>
<td>Control Group</td>
<td>87</td>
<td>91</td>
<td>+4</td>
<td>ns</td>
</tr>
</tbody>
</table>

when specific subscales are analyzed several show significant differences (see tables 6 and 7). Several of the subscales on the ITERS and Arnett reached statistical significance with positive changes in routines (greeting/departing, meals/snacks, nap time, diapering/toileting, health/safety practice/policy) learning activities (eye-hand coordination, active physical play, blocks, pretend play, cultural awareness), sensitivity, and appropriate discipline for the mentoring group. The only statistically significant finding with the control group was in a negative change in interactions in which the scores decreased from pre-test to post-test. Paired t-tests were used in all of these analyses for Tables 6 and 7.

Table 6
Mentoring Group

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITERS subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routines</td>
<td>36</td>
<td>41</td>
<td>.005</td>
</tr>
<tr>
<td>Listening activities</td>
<td>8</td>
<td>9</td>
<td>ns</td>
</tr>
<tr>
<td>Learning activities</td>
<td>28</td>
<td>31</td>
<td>.05</td>
</tr>
<tr>
<td>Interactions</td>
<td>13</td>
<td>13</td>
<td>ns</td>
</tr>
<tr>
<td>Adult needs</td>
<td>17</td>
<td>19</td>
<td>ns</td>
</tr>
<tr>
<td>Arnett subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>26</td>
<td>31</td>
<td>.001</td>
</tr>
<tr>
<td>Appropriate discipline</td>
<td>7</td>
<td>9</td>
<td>.05</td>
</tr>
</tbody>
</table>
Table 7
Control Group

<table>
<thead>
<tr>
<th></th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITERS subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routines</td>
<td>41</td>
<td>42</td>
<td>ns</td>
</tr>
<tr>
<td>Listening activities</td>
<td>9</td>
<td>8</td>
<td>ns</td>
</tr>
<tr>
<td>Learning activities</td>
<td>29</td>
<td>31</td>
<td>ns</td>
</tr>
<tr>
<td>Interactions</td>
<td>15</td>
<td>13</td>
<td>.02</td>
</tr>
<tr>
<td>Adult needs</td>
<td>17</td>
<td>17</td>
<td>ns</td>
</tr>
<tr>
<td>Arnett subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>28</td>
<td>31</td>
<td>ns</td>
</tr>
<tr>
<td>Appropriate discipline</td>
<td>6</td>
<td>7</td>
<td>ns</td>
</tr>
</tbody>
</table>

Discussion

These data demonstrate that the sites that were mentored improved on the ITERS and the Arnett. This is an encouraging result in that the intervention was only 4 months long. It is an important finding because the majority of mentoring projects in the past have utilized anecdotal evidence to demonstrate their effectiveness. Very few programs have conducted randomized trials of their interventions.

It is clear from the data that training/technical assistance interventions are needed in infant toddler programs because of the low scores on various program quality measures. It is also discouraging in that the control programs did not improve in which the ITERS went from 137 (pre-test) to 132 (post-test). This is a finding that will be monitored over time to see if this trend continues. Hopefully this was just an aberration in the data; however there does seem to be support when these data are compared to other studies (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997).

The public policy implications are that an intensive mentoring intervention of only four months can produce positive, although not statistically significant, changes in the overall quality of child care programs both globally and with caregiver interactions. Previous research (Johnson, 1994) has indicated that increasing the number of hours of training produces more developmentally appropriate behaviors in child care staff. Mentoring fits this model because it is an intensive one on one intervention in which the mentor and protégé are engaged in problem...
solving activities to improve the overall quality of the interactions and environment of the child care program.

References


The Instrument Based Program Monitoring Information System and the Indicator Checklist for Child Care

Richard Fiene
Office of Children Youth and Families
Commonwealth of Pennsylvania

Mark Nixon
Children's Services Monitoring Transfer Consortium, Washington, D.C.

ABSTRACT: The Instrument Based Program Monitoring Information System (IPM) and the Indicator Checklist (IC) are two tools for the state management of child day care services. A methodology for monitoring interviews and site visits to child day care programs is described. An integral feature of IPM is a system of assigning weights to the questions or items so that scores reflect the relative importance of state regulations. An Indicator Checklist is a questionnaire or checklist that contains selected, predictive items from a longer, comprehensive instrument that a state uses to monitor child day care providers' conformance to state day care regulations. An Indicator Checklist contains items that have been determined to be most effective in discriminating between providers that typically receive high overall scores on the comprehensive instrument and providers that typically receive low overall scores.

For nearly half a century, state governments have accepted responsibility for ensuring that those who care for children in their home and in day care centers meet minimum requirements for health and safety. During the past decade as the amount of state and federal funds for day care have grown, states have taken an active role in monitoring (1) the ways in which day care providers administer their programs, and (2) the quality of the services provided to children for whose care the state is paying.

Nationally, day care is big business. It is estimated that currently there are more than 118,000 licensed providers who serve an estimated 1.2 million children every day. The stakes in assuring that these children are well served are high, both in terms of public health and safety and from the viewpoint of enhancing the growth and development of America's most precious resource, its children. It is estimated that $6.3 billion dollars are spent annually on day care services.1

Reprints should be requested from Richard Fiene, Directory of Research and Information Systems, Office of Children, Youth, and Families, 1514 North Second Street, Harrisburg, PA 17102.

1 Day care services include group day care centers serving 12 or more children, group day care homes serving 6-11 children, and family day care homes serving 5 or fewer children. Head Start & nursery school programs that operate for part day are included in day care services definition.
However, in monitoring these services, states spend less than one percent of their day care funds each year to ensure that providers comply with regulations or meet quality guidelines.

This article describes an approach in monitoring child day care services called: Instrument Based Program Monitoring (IPM). An IPM differs substantially from the more common approach to monitoring: narrative site visit reports used by most states. The narrative report approach usually includes a site visit to each provider and the preparation of a summary of observations and interpretive and evaluative comments about the monitor's findings. These reports are time consuming to prepare, and often difficult to summarize succinctly for policy makers and administrators. This article describes an alternative to the narrative site report.

**Forces Changing the Regulatory Environment**

The job of state agencies in program monitoring is currently changing in response to powerful forces in American society, especially at the level of state government.

*First*, there is the continuing need to assure parents that their children will not be subjected to unsafe day care environments and that day care providers who receive state funds are meeting the terms of their contracts with the state by providing quality services. Quality services are defined as day care services that promote sound child development principles and do not only ensure that children are in healthy and safe child care environments. Public accountability requires that the state entertain a dual purpose, one is to monitor compliance with state regulations; but secondly and equally important, there is a strong need for the state to ensure that quality child development services are supported and provided.

Gwen Morgan's (1980) work is particularly helpful in providing direction regarding the relationship between licensing and funding criteria. A Model presented by Morgan (1980) clearly delineates a regulatory continuum where day care licensing is considered as the floor to quality with accreditation as the standard of quality for which model day care programs strive. Recent efforts by the National Association for the Education of Young Children (Center Accreditation Project (1983)) and the Children's Services Monitoring Consortium (Child Development Program Evaluation Scale (1984)) have helped to support this move towards accreditation and the measurement of quality in early childhood programs. These efforts take on additional meaning given the direction from the federal government to pass as much of the responsibility for monitoring early childhood programs to the states.
Second, the fiscal cutbacks that are now occurring in many states will almost certainly increase the pressure on state agencies to operate as efficiently as possible. Cutbacks in staff across agencies are likely, even as workloads increase. These factors will force states to streamline their regulatory enforcement and monitoring efforts in all areas, including day care and children's services. A promising approach attempted in some states is moving from a licensing to a registration system. In a registration system, the locus of control for the regulatory process is shifted from the state to the provider level—the provider is responsible for assuring that s/he meets all registration requirements.

Third, the role of the state in regulating private sector organizations is changing. There are now active pressures to reduce the general level of state regulation with a view toward encouraging private market forces in the production and allocation of goods and services. Further, there is a commitment in a growing number of states to reduce the extent of the Federal Government's involvement, including federal funding and accompanying regulatory requirements, in several areas, notably human services (The moratorium placed on the Federal Interagency Day Care Requirements is a specific example which was supported by a number of states).

Fourth, many states are actively seeking ways to reduce the burden on the private sector of the compliance monitoring activities that are performed by the state. For those regulations that continue in force, many states will be examining approaches that simplify monitoring procedures and make them less onerous for providers. This is particularly true for day care services, which are often provided by individuals or organizations that may have little experience coping with regulations.

IPM as a Response to These Forces

One approach that states have used to cope with these forces is the development of Instrument-Based Program Monitoring Systems—(IPMs).

As the name implies, an IPM system incorporates three distinguishing characteristics: First, it is instrument-based. The system uses checklists or questionnaires that contain highly specific questions. These questions usually correspond directly to the state's regulations or other requirements (e.g., fiscal requirements). Second, it supports program monitoring. In its broadest sense, program monitoring is the management process of conducting periodic reviews
or inspections to ensure that certain activities, such as the provision of
day care service, meet acceptable criteria, and the process of effecting
corrective action where required. Program monitoring may include one
or some combination of:

1. Licensing reviews (Table 1 gives a listing of items taken from
Pennsylvania's IPM at the licensing and minimal standards
level);
2. Contract compliance reviews; and
3. Evaluations of program quality that go beyond minimum re-
quirements to health and safety. A specific example that may be
helpful is taken from the California Child Development Program
Quality Review (1982) Instrument. What follows is a sampling
of the Table of Contents:

   PROGRAM QUALITY SUB SCALE
   A. GOALS AND OBJECTIVES OF CHILD DEVELOPMENT
      PROGRAM ARE EVALUATED AT LEAST ANNUALLY
      BY THE STAFF AND PARENTS AND ARE MODIFIED
      AS NEEDED
   B. TEACHING STAFF HIGHLIGHTS EACH CHILD BY
      SHARING INDIVIDUAL ETHNIC AND CULTURAL
      BACKGROUNDS—EMPHASIS IS PLACED ON CARE-
      GIVER OBSERVATIONS.
   C. THE GOALS, OBJECTIVES, AND PROCEDURE FOR
      IDENTIFICATION OF CHILDREN'S NEEDS ARE
      EVALUATED AT LEAST ANNUALLY BY STAFF AND
      PARENTS (Fiene, 1984).

Third, IPM is a comprehensive system. It is part of a group of related
steps such as on-site reviews, corrective action, follow-up reviews, and
summarizing and reporting results that are used recurrently to ac-
complish the task of compliance monitoring. Program, fiscal, and
statistical components can be linked quantitatively to constitute a
comprehensive IPM system for day care. A new software decision sup-
port system (Watson, Fiene, & Woods, 1984) based on IPM is being
developed for micro-computer technology and is being pilot tested in
Michigan Department of Social Services, and Texas Department of
Human Resources. When the IPM system is used in this linked
fashion, it provides the basis for monitoring child day care Vendor &
Voucher Delivery systems.

The advantages of an IPM system that are responsive to the
changes mentioned earlier include: consistency, coverage of all
regulatory areas, clear expectations simplified monitoring procedures,
<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Pennsylvania Child Development Program Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specific Items Within Identified General Areas</td>
</tr>
<tr>
<td>General Requirements</td>
<td>4. Child abuse reporting procedures</td>
</tr>
<tr>
<td>1. Relevant approvals</td>
<td>5. Provision for special services</td>
</tr>
<tr>
<td>2. Insurance coverage</td>
<td></td>
</tr>
<tr>
<td>3. Parent participation</td>
<td></td>
</tr>
<tr>
<td>Staffing Standards</td>
<td>staff requirements</td>
</tr>
<tr>
<td>1. Qualifications of staff</td>
<td>4. Staff health requirements</td>
</tr>
<tr>
<td>2. Responsibilities</td>
<td></td>
</tr>
<tr>
<td>3. Adult/child ratio and minimum</td>
<td></td>
</tr>
<tr>
<td>Employee Records</td>
<td></td>
</tr>
<tr>
<td>1. Evidence of qualifications and references for staff</td>
<td></td>
</tr>
<tr>
<td>Building &amp; Site</td>
<td></td>
</tr>
<tr>
<td>1. Appropriate indoor and outdoor materials</td>
<td></td>
</tr>
<tr>
<td>square footage per child</td>
<td>5. Cleanliness</td>
</tr>
<tr>
<td>2. Characteristics of play areas</td>
<td>6. Screening of windows and doors</td>
</tr>
<tr>
<td>4. Storage of medicine and</td>
<td>8. Educational materials available</td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
</tr>
<tr>
<td>1. Condition and placement of equipment</td>
<td>2. Swimming regulations</td>
</tr>
<tr>
<td>3. Napping rules</td>
<td></td>
</tr>
<tr>
<td>Program for Children</td>
<td></td>
</tr>
<tr>
<td>1. Evidence of written program plan with developmental activities</td>
<td>special needs children</td>
</tr>
<tr>
<td>2. Discipline</td>
<td>4. Sanitary habits developed</td>
</tr>
<tr>
<td>3. Identification and referral of</td>
<td>5. Infant/toddler stimulation</td>
</tr>
<tr>
<td>Food &amp; Nutrition</td>
<td>6. School-age requirements</td>
</tr>
<tr>
<td>1. Menu requirements</td>
<td>3. Utensils</td>
</tr>
<tr>
<td>2. Infant formula rules</td>
<td>4. Special diet considerations</td>
</tr>
<tr>
<td>Transportation</td>
<td></td>
</tr>
<tr>
<td>1. Vehicles all licensed and inspected</td>
<td>4. Restraint of children</td>
</tr>
<tr>
<td>2. Insurance coverage</td>
<td>5. First-aid kit materials</td>
</tr>
<tr>
<td>3. Adult/child ratio</td>
<td></td>
</tr>
<tr>
<td>Child Health</td>
<td></td>
</tr>
<tr>
<td>1. Requirements of health records</td>
<td>4. Medications</td>
</tr>
<tr>
<td>2. Emergency contact information</td>
<td>5. Procedure for ill children</td>
</tr>
<tr>
<td>3. Medical emergency procedures</td>
<td>6. First-aid requirements</td>
</tr>
<tr>
<td>Staff Health</td>
<td>2. Physical requirements for infant caregivers</td>
</tr>
<tr>
<td>1. Procedures for staff illness</td>
<td></td>
</tr>
<tr>
<td>Procedures &amp; Applications</td>
<td></td>
</tr>
<tr>
<td>1. Pre-admission policy</td>
<td>3. Requirements of day care agreement</td>
</tr>
<tr>
<td>2. Requirements for child's application</td>
<td></td>
</tr>
<tr>
<td>Child Records</td>
<td></td>
</tr>
<tr>
<td>1. Frequency of updating records</td>
<td>4. Parental rights to records</td>
</tr>
<tr>
<td>2. Confidentiality</td>
<td>5. Procedure for release of information</td>
</tr>
<tr>
<td>3. Information to be included in child's records</td>
<td>6. Use of records after termination of service</td>
</tr>
</tbody>
</table>
and potential for cost efficiencies. With an IPM system, the same questionnaire or checklist is used with all providers, and there is less opportunity for individual bias in reporting results. Similarly, basing the questions or checklist items explicitly on the regulations or other requirements makes it possible to ensure that all areas are covered adequately. Having a clear set of questions that are known to both monitoring staff and providers reduces the possibility of misunderstandings and misinterpretations concerning the results of the review. Finally, standardized procedures for administering the questionnaire and processing the results can simplify the state’s monitoring task and reduce the time, cost, and burden of monitoring both to the provider and to the state.

Four agencies (Pennsylvania’s Office of Children Youth and Families, West Virginia’s Office of Social Services, California’s Office of Child Development, and New York City’s Agency for Child Development) that are part of a consortium for improving the monitoring of children’s services (Children’s Services Monitoring Transfer Consortium) have experienced significant improvements in provider satisfaction with monitoring efforts and have, in some cases, achieved more efficient allocations of resources for day care and day monitoring. Pennsylvania has experienced substantial cost savings by linking the results of their IPM system to the state’s fiscal and statistical information systems (See Figure 1). The state was able to set a ceiling on

![Pennsylvania Model for Day Care Management-Information-Technical Assistance System](image)
day care funding that did not jeopardize program quality, and used the funds that were formerly given to high-cost providers to improve services of other providers on a targeted basis. The state saved approximately $5 million in day care funds while maintaining the quality of day care services, and it did so without major resistance from the provider groups. California has been able with its IPM system to begin automation of its licensing and program quality instruments and linking these data with unit cost and service information on providers. In the development of the program quality instruments, a representative sample of providers from across the state played a critical role in the development and implementation of California's IPM system. These links are providing the basis for a child development, decision support system for the Office of Child Development in California.

**Indicator Checklist Improves IPM Systems**

Very recently, a number of states (Pennsylvania, West Virginia, Michigan, California, Texas, and New York) have begun experimenting with what has been called an "Indicator Checklist." Simply defined, an indicator checklist is a questionnaire or checklist that contains selected items or indicators from a longer, comprehensive instrument that is used as part of an IPM system. The items on the checklist are those that have been determined to be most effective in discriminating between providers that typically receive high overall scores on the comprehensive instrument or provide a high level of quality care and providers that typically receive low overall scores or provide low level of care (Figure 2).

Because of their value in distinguishing between providers who are in compliance and those that are out of compliance, the items on the in-
indicator checklist have been called “predictor” items. That is, they are a subset of items from the longer instrument that have a strong ability to “predict” the results that would have been obtained had the comprehensive instrument been administered to a given provider. In four of the states mentioned above, the average length of their respective Indicator Checklist’s have been approximately 25 items. This compares with the average of approximately 200 items on their respective comprehensive instruments. The relationship between the scores obtained on the state’s Indicator Checklists and their comprehensive instruments have been extremely high. When a Pearson’s Product Correlation Coefficient was calculated on the Indicator Checklist and the comprehensive instrument for each state the correlation coefficients were always at a $r = +.80$ or higher (See Figure 2a for a graphic display of West Virginia’s data).

**FIGURE 2a**
Based on the results of Pennsylvania’s, West Virginia’s, California’s and New York City’s Indicator Checklists, certain common items were consistently showing up as predictor items that were separating those good providers from those problem providers. In other words, the following items were always in compliance for the good providers and were always out of compliance for the problem providers:

**LICENSING SUBSCALE**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. GROUP SIZE AND ADULT CHILD RATIOS:</td>
<td></td>
</tr>
<tr>
<td>INFANTS</td>
<td>1 STAFF TO 5 CHILDREN</td>
</tr>
<tr>
<td></td>
<td>10 INFANTS IN A GROUP</td>
</tr>
<tr>
<td>TODDLERS</td>
<td>1 STAFF TO 4 CHILDREN</td>
</tr>
<tr>
<td></td>
<td>8 TODDLERS IN A GROUP</td>
</tr>
<tr>
<td>PRESCHOOLERS</td>
<td>1 STAFF TO 10 CHILDREN</td>
</tr>
<tr>
<td></td>
<td>20 PRESCHOOLERS IN A GROUP</td>
</tr>
<tr>
<td>SCHOOL AGE</td>
<td>1 STAFF TO 15 CHILDREN</td>
</tr>
<tr>
<td></td>
<td>30 SCHOOL AGE CHILDREN</td>
</tr>
<tr>
<td>B. SUFFICIENT SPACE—MINIMUM OF 40 SQ FT PER CHILD;</td>
<td></td>
</tr>
<tr>
<td>C. EQUIPMENT IS EASILY ACCESSIBLE TO CHILDREN;</td>
<td></td>
</tr>
<tr>
<td>D. ALL VEHICLES ARE EQUIPPED WITH AGE-APPROPRIATE SAFETY CARRIERS;</td>
<td></td>
</tr>
<tr>
<td>E. CLEANING MATERIALS ARE INACCESSIBLE TO CHILDREN;</td>
<td></td>
</tr>
<tr>
<td>F. EMERGENCY CONTACT INFORMATION IS AVAILABLE FOR ALL CHILDREN;</td>
<td></td>
</tr>
<tr>
<td>G. ALL STAFF HAVE HAD PERIODIC HEALTH APPRAISALS;</td>
<td></td>
</tr>
<tr>
<td>H. ACTIVITIES PROMOTE:</td>
<td>DEVELOPMENT OF SKILLS</td>
</tr>
<tr>
<td></td>
<td>SELF-ESTEEM</td>
</tr>
<tr>
<td></td>
<td>POSITIVE SELF-IDENTITY</td>
</tr>
<tr>
<td></td>
<td>CHOICE OF ACTIVITIES.</td>
</tr>
</tbody>
</table>

To most administrators and policymakers, the advantages of a shorter form will be readily apparent. The short form extends the general advantages of an IPM system in three key ways. 

*First,* it substantially reduces the burden on providers, especially those providers that have a record of high compliance and are judged
suitable for use of the short form—it is proposed that these providers be visited once every three years using the comprehensive instrument. In the intervening years, the indicator checklist should be used.

Second, the indicator checklist approach can further reduce a state’s cost of monitoring and permit the more efficient reallocation of staff resources to other activities. A cost effectiveness study conducted in West Virginia utilizing their indicator checklist resulted in a savings of 50% staff time in determining the level of compliance of providers (in dollars, this translated to $800 annually per visit saved (Peat, Marwick, & Mitchell 1983). With such a substantial savings in time, program monitors/evaluators could be freed to act more as consultants in providing technical assistance to providers.

Third, reviews of providers may be consolidated where appropriate. For example, state staff who perform fiscal/contract compliance audits of providers might be trained to administer the indicator checklist during their audit.

The total effect of maintaining a strong compliance monitoring capability that is less of a burden on providers and that achieves greater efficiency with lower cost is a higher quality monitoring system.

What is Needed to Develop an Indicator Checklist?

An indicator checklist is constructed as follows (See Figure 3):

1) Begin with an existing, comprehensive instrument that has a sufficiently large number of items so as to make greater efficiency desirable. The relative importance of each item as reflected in some kind of scoring or weighting system must have been established. Many criteria may be used for weighting the individual items. One criterion that is particularly useful for weighting purposes is the extent to which a particular item is related to health, safety, or developmental risks to children.

2) Your state should have used the comprehensive instrument long enough so that it is considered reliable for monitoring purposes; the instrument should have generated data that can be used to distinguish among providers in substantial compliance and weak or non-compliant providers.

3) With an existing, comprehensive instrument and some historical score information, it is possible to use a simple arithmetical formula (phi coefficient) to select those items from the long questionnaire that are most useful in distinguishing be-
tween good and inadequate programs. These distinguishing or "predictor" items form the basis of the indicator checklist (See Fiene & Nixon, 1983) for a detailed explanation of the formula for developing an indicator checklist.

4) The final step is to include on the short form particular questions or items from the comprehensive instrument that are of critical importance to the health and safety of children. Typically, these are items which, if violated, would be sufficient basis for denying or revoking a license for a day care program. Usually, such items are few in number. They are added to the short form with the predictor items to ensure that children will not be jeopardized by any statistical errors that might occur if only the "predictor" items were used.

From this description of the procedure for developing the shortened instrument, it is clear that the essential prerequisites for such a checklist are: 1. a long, comprehensive instrument in which state administrators have confidence; 2. items on the comprehensive instrument that are weighted to indicate their relative importance; 3. sufficient score data from use of the comprehensive instrument to differentiate among better and worse programs; and 4. state commitment to developing a short form instrument.

Specific Concerns of Administrators and Policymakers

It may be useful to address particular concerns of administrators and policymakers who may be interested in or even actively considering developing a shortened form of their state's monitoring or

Constructing The Indicator Checklist
licensing questionnaire or checklist. In particular, administrators will need to know: how their state can make use of an indicator checklist; whether indicator checklists have been tried by other states; how the quality of monitoring can be ensured; and whether there are potential drawbacks.

**Can My State Make Use Of An Indicator Checklist?**

Practically every state that presently has some form of questionnaire or checklist can potentially profit from using a shortened form of the instrument. Naturally, if your state’s instrument is already sufficiently short, then little will be gained by being more selective about questions or items to include. Many states are confronted, however, with lengthy instruments that cover a wide range of requirement areas. These states are prime candidates for short-form instruments.

Similarly, perhaps obviously, if your state does not currently have an instrument-based system, then consideration of an indicator checklist/short form is premature.

In order to develop a successful indicator checklist, it is important that the items on your state’s current instrument be clearly linked to:

1. Your state’s requirements (regulations); and
2. The results or outcomes that are considered desireable with respect to the providers’ performance in such areas as licensing, contract monitoring, and program quality.

Unless there is a clear correspondence between instrument items and requirements, there is a danger that the items selected for inclusion on the short form will be only loosely tied to regulations and may be perceived by providers as improper or illegal. Similarly, if there is only a weak link between items on your state’s comprehensive instrument and the results that you expect from providers, then the ground for selecting particular items as good predictors will not be solid enough.

**Have Indicator Checklists Been Tried By Other States?**

The concept of an indicator checklist may be appealing, but administrators are usually hesitant to take risks that could jeopardize systems that have been developed through years of work. It is often satisfying to know that other states have already tested the concept in practice.

At present, the indicator checklist concept is still an innovation that holds great promise but has been fully implemented in only four
states; Pennsylvania, West Virginia, New York, and California have developed an indicator checklist/short form and are testing the concept. Because the initial analyses conducted by these states suggest that the short form can work, other states such as Michigan and Texas have declared their intention to develop a shortened instrument by using these states' experiences as a guide. Clearly though, the indicator checklist/short-form methodology is still in the experimental stage.

**How Can The Quality Of Monitoring Be Ensured?**

Top administrators may wonder whether the shortened instrument presented here will compromise the quality of their state's current monitoring effort. Our view is that the short form will enhance current monitoring efforts by increasing the efficient and effective utilization of monitoring staff. But there are precautions that states should take in developing and using indicator checklists.

The indicator checklist/short instrument should not be used as a substitute for the comprehensive instrument, but rather as its complement. If the short form is viewed as the monitoring instrument, then there may be a tendency over time for providers to meet only the requirements covered on the short form. This situation could, indeed, compromise the quality of monitoring.

On the contrary, we would anticipate that states might keep their comprehensive instruments as the definitive set of compliance expectations and administer them for the initial review (e.g., licensing review) of a provider, and could use the indicator checklist/short form as:

1. A screening device to determine whether, for a given provider, it is necessary to administer the longer version; and
2. An interim review instrument to be used as the principal tool for providers who have a good record of compliance.

For example, the comprehensive instrument might continue to be used for "problem" providers and on a periodic basis, say, every three years for good providers. Naturally, if the short form were used with a provider and problems were discovered, then the comprehensive instrument, or some portions of it, could be administered.

Over time, as conditions change, it will be necessary to update and revise both the comprehensive and short instrument. Using the comprehensive instrument at least periodically with all providers will provide a basis for modifying the short form to reflect changing compliance patterns.
We expect that both versions of the instrument would be used by state staff who are trained and competent to assess compliance. These staff would certainly not limit themselves to using the short form if they determined, on site, that conditions warranted using the comprehensive instrument. The purpose of the indicator checklist/short form is to increase the options available to the state for monitoring in a flexible and cost-effective manner, not to put unreasonable constraints or "blinders" on monitoring staff.

What Are The Potential Drawbacks?

As with all innovations, the introduction of an indicator checklist as the basis for routine monitoring in a state may create some problems. Because so few states have introduced indicator checklists on a widespread basis, it is difficult to identify all of the concerns that may arise in practice. However, a few potential problems can be anticipated. (See Table 2).

First, some states' regulations require that all providers be reviewed every year in all regulatory areas. That is, the state insists that a comprehensive review, for example, using the comprehensive form of a state's monitoring instrument, take place for each provider. If this is the situation in your state, then the use of a shortened instrument may depend on changing the current regulatory provisions concerning the frequency and scope of reviews. A strong basis for making such a change is the cost effectiveness of the indicator checklist/short form, that is, its potential for reducing monitoring costs substantially without reducing the quality of the monitoring effort.

<table>
<thead>
<tr>
<th>Potential Drawbacks</th>
<th>Possible Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory Requirement for Annual Comprehensive Review</td>
<td>Change Regulatory Requirements</td>
</tr>
<tr>
<td>Staff Resistance</td>
<td>Educate Staff</td>
</tr>
<tr>
<td>State's Lack of Prerequisites</td>
<td>Seek Assistance in Obtaining Prerequisites</td>
</tr>
</tbody>
</table>

Second, the state's staff who are responsible for monitoring may resist the introduction of the indicator checklist/short form. From their viewpoint, it may appear that the use of indicator checklists is a reduction in the importance of their professional roles and that the
state's cost savings may take the form of fewer jobs for day care monitors.

In our view, states may need to assure their staff that the indicator checklist/short form is not intended to reduce either the professional judgments involved or the scope of the monitoring function. As mentioned earlier, the comprehensive and short instruments must be used in a complementary way, not as substitutes, in order for the short form to have validity. If anything, the judgment of the monitors may be expanded as it becomes necessary to decide whether, in a particular case, the short instrument will be sufficient to measure compliance with state requirements, and/or program quality criteria. Monitors must be persuaded that the short form is an aid that is designed to reduce the monitors' workload for those providers with whom the short form is appropriate.

The reduction in workload may gradually change the relationship of monitors to providers from one of regulation to one of active support in improving the health and safety of the day care environment and encouraging child development. This change in the monitors' role could enable the state to make even better use of the current monitoring staff's knowledge and experience.

With respect to costs and staff reduction, there is little question that substantial decreases in workload could also result in reduced staffing levels. However, before considering cutbacks in staff, we would encourage states to consider reallocating staff time that is saved because of the short form to other monitoring activities such as technical assistance to providers involving program quality issues.

Third, a state may discover that it does not have the necessary prerequisites, described earlier, to develop and implement an indicator checklist. If your state lacks these prerequisites—in particular a comprehensive instrument, reports of scores, and a system of weighting items on the instrument—then it may be advantageous for you to examine other reports prepared by the Children's Services Monitoring Transfer Consortium that describe how these prerequisites can be met. You may be interested in obtaining the Consortium's series of Guide Books. The three volumes of this series describe in detail how to develop a comprehensive instrument from which an indicator checklist/short form can be derived.

Conclusion

The art of monitoring has evolved considerably in recent years as more highly trained staff have been given responsibility for monitoring, and as clearer procedures, such as instrument-based program monitoring, have been implemented. This evolution has con-
tributed positively to achieving the desirable outcomes of improved day care for children for which the state has developed regulations. At the same time, the evolution has, we hope, made it possible for providers to operate more effectively with the minimum necessary oversight by the state.

Instrument Based Program Monitoring Systems are now being developed in other children’s services such as MH/MR services. Pennsylvania has developed its child welfare information system based on the instrument based program monitoring concept. This system meets two needs for Pennsylvania: it tracks children through its foster care system; and it complies with PL 96-272—the Adoption Assistance and Foster Care Act—a federal law. West Virginia is attempting to use the IPM methodology in monitoring its family day care home programs.

Also, a micro-computer, decision support system based on the Instrument Based Program Monitoring and Indicator Checklist methodology is being developed by the Children’s Services Monitoring Transfer Consortium (CSMTC). The CSMTC is a group of states (Pennsylvania, West Virginia, California, New York, Michigan, and Texas) who have been disseminating exemplary monitoring techniques from state to state. Based on the combined efforts of these states, a generic indicator checklist that measures compliance with state regulations as well as program quality has been developed (Fiene, 1984). The CSMTC feels that this generic indicator checklist can be used by states who have not developed an instrument to assess providers, or as a model instrument to assist states in developing their own instruments.

The real potential of monitoring in achieving social goals, (such as protecting the health and safety of young children, ensuring quality child development programs, and tying these to child development outcomes), will be better realized through continuing research and development of improved monitoring procedures. It is in this context that the development of the indicator checklist represents a major advance in monitoring children’s services.

References


Merrill, B. *West Virginia Indicator Checklist*, West Virginia Department of Social Services, Charleston, W.V., 1981.


Morgan, G. Regulating early childhood programs in the eighties, unpublished manuscript, 1980.


Peat, Marwick, & Mitchell, Assessment of the implementation of the indicator checklist for day care centers in West Virginia, *Children's Services Monitoring Consortium*, September, 1983.


