CONTEMPORARY ISSUES IN LICENSING
Monitoring Strategies for Determining Compliance: Differential Monitoring, Risk Assessment, and Key Indicators

Introduction

While States’ licensing systems primary goal is to improve the health and safety of children in child care, important decisions must be made in order to also maximize administrative cost efficiencies. With limited resources, licensing administrators work to ensure that monitoring visits focus on what is most important in keeping children safe. In the absence of research that assesses the efficacy of various approaches, States are moving ahead with different methods to identify and reduce the risk of harm to children. Some strategies include:

- Identifying licensing rules where violations pose a greater risk to children;
- Assigning a weight to each rule to further distinguish levels of regulatory compliance;
- Focusing monitoring visits on key indicators from the rules that predict compliance and reduce risks;
- Increasing monitoring frequency for programs with low levels of compliance;
- Increasing monitoring depth for programs with low levels of compliance;
- Helping providers, parents, and licensing staff better understand the potential consequences of serious noncompliance;
- Identifying providers in need of technical assistance; and
- Using more sophisticated data systems to target case management and improve consistency in enforcement actions.

The purpose of this report is to describe various methods States are using to monitor child care facilities efficiently and effectively. It provides descriptions and examples of these methods and details of States’ practices.

---

Content | Page
---|---
Introduction | 1
Methodology | 2
Methods for Monitoring Compliance | 2
- Differential Monitoring | 4
- Full and Abbreviated Compliance Reviews | 5
Approaches to Identifying Critical Rules | 6
- Key Indicators | 6
- Risk Assessment | 7
Issues to Consider | 10
References | 11

---

1 Because licensing of child care facilities most often occurs at the state level, there are variances in terminology from State to State. For the purposes of this report, the terms identified are defined as follows and are used interchangeably throughout: Regulations, Rules, Requirements, Policies and Administrative Code, Laws, Statutes.
Methodology

To support the Office of Child Care’s goal of children served in safe, healthy child care settings, the National Center on Child Care Quality Improvement (NCCCQI) contracted with a group of nationally-recognized consultants with expertise in administering and researching licensing systems to prepare a series of written reports about critical licensing issues.

The information provided in these reports was obtained by surveying and interviewing representatives of state licensing agencies in nine States: CT, FL, GA, NC, OH, OK, TX, UT, and WA. The States selected are not a representative sample but were chosen based on the consultants’ knowledge that they are implementing effective and innovative practices which may be helpful to other state licensing agencies. Additionally, an effort was made to achieve some degree of geographic representation through the States selected.

Licensing personnel from the nine States selected first completed a written survey instrument and then spoke with the consultants in a telephone interview. All individuals interviewed were licensing agency directors or top-level administrators.

Information from Research Brief #1: Trends in Child Care Center Licensing Regulations and Policies for 2011 (NCCCQI, 2013) and The 50-State Child Care Licensing Study: 2011-2013 Edition (National Association for Regulatory Administration [NARA], 2013) are also included to provide national data and context to the information gathered from the nine States. Both of these reports include data gleaned from a national survey of licensing agencies conducted by NARA. Responses to the NARA survey were received from licensing agencies in all 50 States and the District of Columbia.²

Methods for Monitoring for Compliance

In an effort to ensure the health and safety of children in child care facilities, States seek to identify and assess the risk of harm to children and increase monitoring in programs with lower levels of compliance. At the same time, state licensing agencies need to make the most efficient and effective use of available, and often shrinking, resources.

NARA, in Recommended Best Practices for Human Care Regulatory Agencies (2009), presents the characteristics of a strong licensing agency, including:

Maintains a research-based risk-assessment method whereby industry-wide and facility-specific risks, including both immediate and cumulative risks, are identified and prioritized; focuses inspections and technical assistance accordingly; and, applies the agency’s enforcement continuum systematically to avert or abate priority risks, to build consistent compliance, and to improve overall consumer protection across all relevant domains. (p. 6)

² In the NCCCQI and NARA reports, and in this report, the District of Columbia is included in state counts and not listed separately.
There are a variety of methods that many States are using, often in combination, in their monitoring and enforcement of licensing rules and regulations. This report explores these methods:

- **Differential Monitoring**: A regulatory method for determining the frequency or depth of monitoring based on an assessment of a facility’s history of compliance with rules;
- **Full and Abbreviated Compliance Reviews**: Conducting an inspection by monitoring all rules (full review) or a selected set of rules (abbreviated review);
- **Risk Assessment**: An approach that focuses on identifying and monitoring those rules that place children at greater risk of mortality or morbidity if violations or citations occur; and
- **Key Indicators**: An approach that focuses on identifying and monitoring those rules that statistically predict compliance with all the rules.

The relationship among these methodologies is often confused, partly because of varying definitions. The graphic below explains the relationship of the methodologies with differential monitoring as the overarching approach and risk assessment and key indicators as types of abbreviated compliance.
**Differential Monitoring**

Differential monitoring is a regulatory method for determining the frequency or depth of monitoring based on an assessment of a facility’s history of compliance with licensing rules. A differential monitoring system can be used to recognize a provider’s strong record of licensing compliance with abbreviated or less frequent inspections if there have been no serious violations for a period of time. For providers with rule violations and compliance issues, licensing agencies can use differential monitoring to focus more attention on those facilities with additional monitoring visits, targeting visits on the problem areas, and providing technical assistance. When inspections are focused on a subset of rules, States often have an option for licensing staff to conduct a full review when necessary.

In its analysis of licensing trends, NCCQI (2013) noted that more than 50 percent of States report having a method for determining the frequency or depth of monitoring based on an assessment of compliance with regulations. *The number of States using differential monitoring increased significantly from 11 States in 2005 to 26 States in 2011.*

**Why Use Differential Monitoring?**

- Increase monitoring frequency for programs with low levels of compliance;
- Identify providers in need of technical assistance;
- Recognize programs with strong compliance records with abbreviated inspections; and
- Use staff resources efficiently.

**Differential Monitoring Policies in Oklahoma**

Oklahoma increases the number of monitoring visits from the required three annual visits if there is a pattern of noncompliance. Technical assistance is provided during all visits as needed. According to the Oklahoma Department of Human Services’ (2012) policies:

After each monitoring visit, licensing staff enter the monitoring frequency plan... [in] the licensing database. Any changes in the monitoring plan must be reviewed by the supervisor. Examples of the required number of visits include:

- One visit per year for inactive child care centers, part-day, or school-age facilities;
- Two visits per year for part-year programs;
- Three visits per year for facilities with a history of compliance;
- Six visits per year for applications, six-month permits, and changes in facility class except a large FCC home changing to a FCC home; and
- Twelve visits per year for seriously noncompliant facilities.

On occasions when licensing staff visit a facility between monitoring visits for purposes such as picking up paperwork, providing consultation on a specific issue, or verifying a required repair or purchase, a full monitoring visit is not required and the visit is not counted toward the required number of visits. If numerous, repeated or serious noncompliance is observed during the visit, a complete monitoring visit is conducted. If caseloads prevent licensing staff from conducting the required number of monitoring visits, the supervisor consults with the staff on case management, and the number of required visits may be reduced if approved by the regional program manager. This adjustment is approved and documented in the case record by the supervisor. Required visits to nonproblematic licensed facilities may be reduced by one visit per year for no longer than a one-year period. More information about 340:110-1-9. Case Management, Instructions to Staff, is available at [http://www.okdhs.org/library/policy/oac340/110/01/0009000.htm](http://www.okdhs.org/library/policy/oac340/110/01/0009000.htm).
Full and Abbreviated Compliance Reviews

States generally conduct full compliance reviews during monitoring visits where all possible areas of regulatory compliance are measured and every rule is checked for compliance. According to NARA (2013), States typically conduct a full compliance review of programs every 1 – 2 years, most often as part of the license renewal process.

A growing number of States are using an abbreviated compliance review to conduct at least some inspections. NCCQI (2013) reported that more than 55 percent of States in 2011 were using abbreviated compliance reviews for some inspections, mostly during routine compliance reviews.

States have different approaches to deciding if and when to use an abbreviated compliance form. NARA (2013) reported that in 2011, most of the States that use abbreviated compliance forms had policies on when to switch from an abbreviated compliance review to a full compliance review. The following examples illustrate how States determine when to use full and abbreviated compliance reviews:

- **Florida** inspects centers a minimum of three times per year and family child care (FCC) homes two times per year. As part of the 1996 WAGES Act, the Florida Legislature directed the Department of Children and Families and local licensing agencies to develop and implement an abbreviated inspection plan for child care facilities based on certain statutory criteria. Florida has an automated child care inspection system that tracks violation data and identifies the providers eligible for abbreviated inspections. Eligible providers have had no Class I or Class II (most serious) violations for two consecutive years. If violations are found during an abbreviated visit, the provider is no longer eligible to be monitored using the shorter form and must have a full compliance review. Florida’s laws about conducting abbreviated inspections are available in 2013 Florida Statutes Sections 402.26 – 402.319 Child Care at http://nrckids.org/default/assets/File/StateRegs/FL/FL_Statutes_402_26-402_319_Child_Care.pdf.

- **Georgia** licensing agency conducts a minimum of two visits per year, including a licensing study and a monitoring visit. The licensing study is a full inspection using an inspection form that includes all rules, with the core rules highlighted (see page 8 for more detail). Monitoring visits involve the use of an abbreviated form that only includes the core rules.

- **North Carolina**’s state statute requires that all providers are inspected by the licensing agency at least once per year, in addition to annual inspections by local or state health and fire inspection agencies. For programs to receive an abbreviated monitoring visit, they must have a four- or five-star license and a compliance score of 85 percent over the past 18 months prior to the scheduled visit date. In the rated license system, higher star levels are obtained by meeting additional requirements related to program quality standards and education levels of staff. “Chapter 110 Child Care Facilities,” in North Carolina General Statutes (2013) is available at http://nrckids.org/default/assets/File/StateRegs/NC/07-13%20Article%2007.pdf.

- **Texas** inspectors and investigators determine which standards to evaluate prior to the inspection but have the ability to add standards during the inspection, if needed. All standards must be evaluated at least once every two years. Standards may be re-evaluated as a result of investigations, follow up on previous deficiencies, or as part of a corrective action. Texas policies on Preparing for Inspections are in Section 4140 in the Texas’ Licensing Policy and Procedures Handbook at http://www.dfps.state.tx.us/handbooks/Licensing/Files/LPPH_pg_4000.asp#LPPH_4140.

- **Utah** inspects centers and FCC homes twice a year. All providers receive an abbreviated unannounced compliance review and a full announced compliance review annually. All of Utah’s announced (full) and unannounced (abbreviated) inspection checklists are available on its Web site at http://health.utah.gov/licensing/centerinspectionchecklists.htm (centers) and at http://health.utah.gov/licensing/HomeInspectionChecklists.htm (FCC homes).
Approaches to Identifying Critical Rules

Often differential monitoring involves monitoring programs using a subset of the licensing rules to determine compliance. There are two methods that States have used to identify these critical rules:

- **Key Indicators**: An approach that focuses on identifying and monitoring those rules that statistically predict compliance with all the rules; and

- **Risk Assessment**: An approach that focuses on identifying and monitoring those rules that place children at greater risk of mortality or morbidity if violations or citations occur.

Focusing on specific rules, whether through a key indicator or risk assessment process or a combination of both, can assist the licensing agency to:

- Implement a differential monitoring policy;
- Guide case management such as targeted technical assistance or witnessed visits;
- Determine enforcement actions based on categories of violations; and
- Assist families in better understanding the potential impact of noncompliance on their child’s care.

**Key Indicators**

Here we describe the key indicator approach, where States identify those rules that statistically predict overall compliance. A methodology for key indicators was developed by Dr. Richard Fiene at Pennsylvania State University. Dr. Fiene (2014) states that “if a program is 100% in compliance with the Key Indicators, the program will also be in substantial to full compliance with all rules. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators, the program will also have other areas of non-compliance with all the rules.” (p. 3)

The indicator methodology was based on research to study the impact of child care quality on children’s development and the relationship between program quality and compliance with state licensing rules (Fiene, 2013). Several conducted in Pennsylvania in the 1980s found that programs in substantial compliance with licensing rules had better quality than those with 100% compliance (with a focus on recordkeeping), which led to including more program items in licensing rules. The studies supported greater use of indicators to save monitoring time and permit more technical assistance and consultation on quality improvement (Fiene, 1986, Kontos & Fiene, 1987).

The key indicators approach is often used to determine the rules to include in an abbreviated inspection form or checklist. Some States have worked with Dr. Fiene to implement his statistical methodology; however, other States have determined indicators by reviewing their rules and choosing by consensus those considered most critical to protecting children’s health and safety. In addition, States that use key indicators often include a few additional rules in their inspections, based on level of risk or random selection.
Washington Employs Key Indicator System

Washington based its system of monitoring checklists on the thirteen indicators developed by Dr. Richard Fiene (2002) for the U.S. Department of Health and Human Services a number of years ago. These are used across all types of programs—centers, FCC homes, and school-age programs. Providers with nonexpiring full licenses are monitored using an abbreviated checklist when the site has demonstrated a high level of compliance since the prior visit. This includes, but is not limited to, no valid complaints, compliance agreements, or other information demonstrating noncompliance with licensing rules. Licensors are required by policy to move to a full checklist in cases where providers are not in compliance with any of the key indicators. Washington has started to use electronic licensing forms and data gathering that will allow for statistical weighting in future years, after the data have matured. Washington’s licensing agency includes some rules in addition to the key indicators in their abbreviated checklists.

- **Policies and Procedures**

- **Monitoring Tools**

**Risk Assessment**

A risk assessment approach can be used to determine the rules that pose a greater risk of harm to children if violated. Risk assessment is most often tied to classifying or categorizing rule violations and can be used to identify rules where violations pose a greater risk to children, distinguish levels of regulatory compliance, or determine enforcement actions based on categories of violations.

There are a number of ways licensing regulations can be assessed for risk, including the following:

- Probability of harm (high, medium, low);
- Severity of harm (extreme, serious, moderate, low); or
- Frequency of violation (numerous, repeated) based on those considered most critical to protecting children’s health and safety.

States that choose a risk assessment approach must determine whether to assign a risk category to all rules or a selected set of rules. A risk category might be assigned to all rules so that enforcement can be tied to level of risk. For example, Florida has categorized all rule violations based on the threat of harm to children:
“Class I Violations” are the most serious in nature, pose an imminent threat to a child including abuse or neglect, and could or do result in death or serious harm to the health, safety, or well-being of a child;

“Class II Violations” are less serious in nature than Class I violations, and could be anticipated to pose a threat to the health, safety, or well-being of a child, although the threat is not imminent; and

“Class III Violations” are less serious in nature than either Class I or Class II violations, and pose a low potential for harm to children.

Alternatively, a licensing agency might only assign a risk category to a subset of rules if the primary purpose of risk assessment is to determine the need for further monitoring visits. **Ohio** has defined Serious Risk Noncompliances (SRNC) for centers and group child care homes based on requirements with the highest risk of harm if violated. Regulations are organized into three large categories: 1) Lack of Supervision, 2) Administrative Negligence, and 3) Environmental Hazards. If a program has a certain number of serious risk violations, they receive additional full compliance inspections. A summary document of the requirements chosen as SRNC is available at [http://jfs.ohio.gov/cdc/RiskRules.pdf](http://jfs.ohio.gov/cdc/RiskRules.pdf).

**Oklahoma** has also identified serious noncompliances that expose children to conditions that present an imminent risk of harm. Their policies clarify that “Imminent risk of harm must be assessed based on the age of the child, the amount of time the caregiver was out of compliance, and the caregiver’s efforts to mitigate the risk. Serious noncompliances are identified through licensing observations, confirmed complaint investigations, and/or self-reported incidences.” The policies are available at [http://www.okdhs.org/library/policy/oac340/110/01/0009003.htm](http://www.okdhs.org/library/policy/oac340/110/01/0009003.htm).

Some States use risk assessment to classify violations and determine enforcement approaches. For example, in **Florida**, violations of the minimum health and safety standards are automatically classified as Class I, Class II, or Class III based on the potential for harm to a child. Enforcement actions, such as monetary fines, are determined by the classification of violations and number of occurrences in a progressive enforcement model. Licensing inspection reports are posted on the Florida Department for Children and Families Web site and include violation classifications. The definitions of the three classes are found in “Chapter 65C-22 Child Care Standards” of the Florida Administrative Code (8/1/2013) at [http://nrckids.org/default/assets/File/StateRegs/FL/FL_Chapter_65C-22.pdf](http://nrckids.org/default/assets/File/StateRegs/FL/FL_Chapter_65C-22.pdf).

In **Utah**, rule violations are classified as Level 1, 2, or 3 violations, depending on both the seriousness of harm to a child that could result from the violation, as well as the likelihood that harm will occur. Level 1 findings are categorized as “cited” findings the first time they occur. Level 2 and 3 findings are initially classified as “technical assistance” findings, which mean that providers are given technical assistance and the opportunity to correct the violation. The number of rule violations and the severity of the violations determine if providers may be placed on a conditional license with additional monitoring inspections. The frequency of monitoring inspections may also increase due to noncompliance during the conditional period. Definitions of the violation levels are available in the “Introduction” section of the *Child Care Center Rule Interpretation Manual* at [http://health.utah.gov/licensing/rules/Interpretation/Center/Section%201-%20Introduction.pdf](http://health.utah.gov/licensing/rules/Interpretation/Center/Section%201-%20Introduction.pdf). Utah’s interpretation manuals include noncompliance levels for each licensing requirement. The manuals are available at [http://health.utah.gov/licensing/rules.htm](http://health.utah.gov/licensing/rules.htm).
In Texas, all of the Child Care Licensing Minimum Standards have been assigned a weight (High, Medium High, Medium, Medium Low, or Low) based on the risk that a violation of that standard presents to children. Weights are noted within the minimum standards documents in the left margin next to each standard or subsection. Only those standards that can be violated (marked as a deficiency) are weighted. The weighted enforcement system utilizes the program’s operations compliance history including the repetition of violations, investigations, types, and number and weight of deficiencies to generate the enforcement recommendations. The Texas licensing standards are available at http://nrckids.org/index.cfm/resources/state-licensing-and-regulation-information/texas-regulations/.

In Texas, inspectors and investigators determine which standards to evaluate prior to the inspection but have the ability to add standards during the inspection, if needed. All standards must be evaluated at least once every two years. Standards may be re-evaluated as a result of investigations, follow up on previous deficiencies, or as part of a corrective action. The weighted enforcement system utilizes the operations compliance history, including the repetition of violations, investigations, types, and number and weight of deficiencies to generate the enforcement recommendations.

Licensing staff document observations to capture the scope and severity of the deficiency, but the weighted standards are now part of the licensing database and decisionmaking process, resulting in more consistent and equitable enforcement practices. The Child Care Licensing Automation Support System (CLASS)* Risk Review is a tool that supplements the professional assessments of licensing staff. The CLASS Risk Review produces enforcement recommendations based upon the type, number, weight, and repetition of violations over the course of an operation’s two-year compliance history. A Risk Analysis summary can be requested by staff seeking feedback on corrective actions. Facilities with serious deficiencies or a significant number of deficiencies, repeat deficiencies, or that fail to make timely corrections, are inspected more frequently to monitor the level of risk to children.

For more information, see “Section 4500: Evaluating Risk to Children” in the Texas Licensing Policy and Procedures Handbook at http://www.dfps.state.tx.us/handbooks/Licensing/Files/LPPH_pg_4300.asp#LPPH_4500.

*CLASS is the Child Care Licensing Automation Support System. It is a computer application used by Texas licensing staff for record management.
Georgia’s Core Rules

Georgia uses a core rule reference chart to determine and assess the health and safety risk of noncompliance to children. When child care licensing consultants conduct inspections, they use the chart to assess the level of severity of the violation and guide their decisionmaking on issuing citations. Each time any core rule within the core rule categories is cited, the risk level of the citation is also assessed. Risk level is assigned at low, medium, high, and extreme levels. The number of core rule categories cited and the assigned risk level determines the annual compliance level. A facility’s annual compliance status is determined on June 30 of each year, based on the performance for the past fiscal year (July 1-June 30), is posted on the public Web site, and remains in place for the next fiscal year. Additional information about Georgia’s core rules is available at http://decal.ga.gov/ChildCareServices/CoreRulesInformation.aspx.

Family Day Care Home Rule Categories
- Criminal Records Check
- Discipline
- Field Trips
- Infant Sleeping Safety Requirements
- Overcrowding Registration Requirements
- Physical Plant
- Playgrounds
- Staff:Child Ratios
- Supervision
- Swimming Pools and Water Related Activities

Child Care Learning Center and Group Day Care Home Core Rule Categories
- Diapering Areas and Practices
- Discipline
- Field Trips
- Infant Sleeping Safety Requirements
- Hygiene
- Medications
- Physical Plant
- Playgrounds
- Staff:Child Ratios
- Supervision
- Swimming Pools and Water Related Activities
- Transportation

Issues To Consider

The goal of differential monitoring, abbreviated compliance tools, risk assessment and key indicators is to create efficiencies and greater effectiveness in monitoring and enforcement, permitting more time for monitoring, especially of those facilities with lower compliance that need more technical assistance and program consultation. It should be noted, however, that these strategies should only be implemented when built on a strong licensing structure with a foundation of adequate periodic unannounced inspections. The States surveyed for this report use different tools and methodologies for measuring compliance, and feel that this practice has increased their enforcement capability. The increased use of these methodologies across States raises some questions for the field to consider:

- While abbreviated compliance forms are widely used, most are not developed using a methodology that statistically predicts compliance. Are all of these methods equally effective in measuring the level of compliance with licensing rules?
- Are all abbreviated compliance systems successful in creating both efficient and effective use of resources? What are the similarities and differences and what is their impact on effective regulation?
- What is the best mix of the measurement methodologies discussed in this report for consistent and strong enforcement of the licensing rules?
How do these methods impact the relationship between licensing and other entities that monitor child care programs, such as Head Start, Quality Rating and Improvement Systems, prekindergarten, and national accreditation?

States must continue to educate providers on the importance of meeting all licensing rules, not only those that are identified as being critical to children’s health and safety. Licensing staff should receive training and guidance on remaining diligent during all on-site inspections, and carefully observing and assessing all facets of the physical facility and program including interaction between staff and children. Licensing policy and procedures should also guide staff on what factors will trigger a full compliance review at any inspection using an abbreviated tool. Lastly, research is needed to compare the various forms of abbreviated compliance systems for their effectiveness in measuring compliance levels and fostering improved compliance and quality.

References


VALIDATION OF QUALITY RATING AND IMPROVEMENT SYSTEMS FOR EARLY CARE AND EDUCATION AND SCHOOL-AGE CARE
DISCLAIMER:
The views expressed in this publication do not necessarily represent the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families or the U.S. Department of Health and Human Services.

ACKNOWLEDGMENTS
The authors would like to thank Ivelisse Martinez-Beck and Naomi Goldstein at the Office of Planning, Research and Evaluation, Kathryn Tout at Child Trends, and Laura Hamilton at RAND for their guidance and feedback on this paper.
Validation of Quality Rating and Improvement Systems for Early Care and Education and School-age Care

Research-to-Policy, Research-to-Practice Brief  OPRE2012-29

April 2012

Submitted to:
Ivelisse Martinez-Beck, PhD., Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Submitted by:
Gail L. Zellman, RAND Corporation
Richard Fiene, Pennsylvania State University

Contract Number: GS10F0030R
Project Director: Kathryn Tout
Child Trends
4301 Connecticut Ave NW
Washington DC, 20008

Suggested Citation:

This Brief was developed by members of the Quality Initiatives Research and Evaluation Consortium (INQUIRE) which is designed to facilitate the identification of issues and the development and exchange of information and resources related to research and evaluation of quality rating and improvement systems (QRIS) and other quality initiatives. INQUIRE is funded by the Office of Planning, Research and Evaluation through the Child Care and Early Education Policy and Research Analysis and Technical Expertise contract with Child Trends.
Validation of Quality Rating and Improvement Systems for Early Care and Education and School-age Care

Quality Rating and Improvement Systems (QRIS) for early care and education and school-age care programs are designed to collect information about quality and to use that information to produce program-level ratings, which are the foundation of a QRIS. The ratings are intended to make program quality transparent for parents and other stakeholders and to encourage the selection of higher-quality programs. The ratings also provide benchmarks that can support efforts to help programs improve their quality. Validation of a QRIS is a multi-step process that assesses the degree to which design decisions about program quality standards and measurement strategies are resulting in accurate and meaningful ratings. Validation of a QRIS provides designers, administrators, and stakeholders with crucial data about how well the architecture of the system is functioning. A carefully designed plan for ongoing validation creates a climate that supports continuous quality improvement at both the program and system level.

To date, QRIS validation efforts have been limited. One reason may be that validation is a complex endeavor that involves a range of activities. In addition, there has been little guidance available that clarifies the purpose of QRIS validation or identifies the activities that comprise validation. At the same time, there is growing pressure to validate these systems as stakeholders seek evidence that QRIS are functioning as intended. The federal government has elevated QRIS validation by including it as a central component of the 2011 Race to the Top Early Learning Challenge and requiring state applicants to develop QRIS validation plans as part of their submissions.

The purpose of this Brief is to help QRIS stakeholders better understand validation and to outline a set of complementary validation activities. The Brief defines validation, describes different types of validation studies, and provides guidance on developing a validation plan, including tools to determine the appropriate scope and timing of validation activities. It also lists references and resources for those who wish to learn more. This Brief is aimed at readers in positions to authorize, finance, design, and refine QRISs and other quality improvement efforts, including state child care administrators, early education policy and program specialists, legislators, and other potential funders.
QRIS Validation and Its Role in Continuous System Improvement

*Validation* is a multi-step process that assesses the degree to which design decisions about QRIS program quality standards and measurement strategies are resulting in accurate and meaningful program ratings.¹ Validation is particularly important for QRISs because these systems at their core rely on ratings of program quality. They are built on the assumption that the quality of early childhood and school-age programs can be reliably measured and that differences in quality across these programs can be identified through the use of a set of quality indicators. Validity data can support conclusions about whether such quality indicators measure quality well and whether the strategies used to combine measures and develop ratings are working as intended (Cizek, 2007).² Valid ratings are critical to QRISs because parents and other stakeholders use these ratings to select the highest-quality care that they can afford. The overall quality rating also carries increasingly high stakes for programs. Indeed, the theory underlying QRISs intentionally creates those stakes to motivate both provider and parent behaviors in support of increased quality (e.g., Zellman et al., 2008; Zellman et al., 2011). In addition to attracting more children, programs that score well may receive higher subsidies for subsidy-eligible children, and may qualify for grants, incentives, and tax credits.

Validity is not determined by a single study; instead, validation should be viewed as a continuous process with multiple goals: refining the ratings, improving system functioning, and increasing the credibility and value of rating outcomes and of the QRIS system as a whole. A carefully designed validation plan will 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). Ongoing validation activities that are carried out in tandem with QRIS monitoring activities (that aim to examine ongoing implementation of the QRIS) and evaluation activities (that examine the outcomes of QRIS) can help a QRIS improve its measures and effectiveness throughout its development and implementation (see Lugo-Gil et al., 2011 and Zellman et al., 2011 for guidance on developing a comprehensive QRIS evaluation).

---

¹ The definition of validation has changed over time. Rather than identifying separate types of validity (construct, predictive, face, concurrent, and content), the current notion is that construct validity includes all evidence for validity, including content and criterion evidence, reliability, and the wide range of methods associated with theory testing (Messick, 1975, 1980; Tenopyr, 1977; Guion, 1977; Embretson, 1983; Anastasi, 1986). As a consequence, we do not differentiate types of validity in this brief.

² Reliability represents the ability of a measure to assess its target behaviors or characteristics consistently. In the case of QRISs, reliability refers to the extent to which independent raters produce similar ratings on individual QRIS elements and on the summary rating (inter-rater reliability) as well as the degree to which raters are consistent over time in their ratings (intra-rater reliability). Such consistency is a prerequisite for validity of any measure.
QRIS validation activities may produce three important benefits. First, validation evidence can promote increased support for the system among parents, ECE-SAC providers and other key stakeholders. Ratings that match the experiences of parents and providers can build trust in the ratings and increase the overall credibility of the system. Second, a system that is measuring quality accurately is better able to target limited quality improvement supports to those programs and program elements most in need of improvement. 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 or that is not measuring quality accurately, it can be removed or revised. For example, measures that vary little if at all across providers whose quality varies substantially in other ways make little or no contribution to quality ratings. Measures of family engagement that include parent ratings are particularly prone to this problem, as parents who have chosen to use and continue to rely on a given provider are highly likely to see the care as good and to rate it according to their views (Zellman and Perlman, 2006; McGrath, 2007; Keyes, 2002; Kontos et al., 1987; Shimoni, 1992). If all or almost all programs receive high ratings on the family engagement measure, then that component of the rating may not be working to distinguish between lower-quality and higher-quality programs. It may be considered important to collect measures of family engagement to ensure that providers continue to focus on it. But knowing that a given measure is not contributing to an overall program quality rating may motivate program developers to consider another way to measure the concept, which might both increase the value of the measure and reduce measurement costs. Indeed, understanding the relationships among rating elements through validation studies can save substantial time and effort.

Despite the importance of validation activities to strengthen QRIS, support for these activities may be impeded by limited resources and concern about the value of validation activities. In states with more mature QRISs, there may be reluctance among stakeholders to assess an established system. In newer systems, policymakers may question the need for validation given the arguments recently offered in support of establishing the system. Validation plans can address each of these concerns by providing evidence to help the system run more efficiently and to establish a climate of continuous improvement. A validation plan will clarify that the system is open to change, intent on improvement, and dedicated to increasing the odds of reaching its goals.

**Designing and Implementing Validation Efforts**

A comprehensive validation plan includes multiple studies that rely on different sources of information and ask different but related questions. These can be understood and organized around four complementary and interrelated approaches to validation. In this section we provide details of the four approaches. Summaries of these details are provided in two tables. Table 1 presents an overview of the four approaches including the purpose of each approach, the activities that might be undertaken, the questions that are asked and the limitations of each approach. Table 2 presents the data needed, data sources, and analysis methods for selected studies within each approach.³

³ The four basic approaches described in the table are very similar to and compatible with those used in the QRIS Evaluation Toolkit (Lugo-Gil et al., 2011).
When reviewing the tables and the remainder of the Brief, it is helpful to be familiar with how three key QRIS terms – component, standard and indicator – are defined. The term quality **component** refers to the broad quality categories used in QRIS (such as staff qualifications, family engagement, and the learning environment). A quality **standard** is defined as a specific feature of quality such as specialized curriculum and assessment training in the staff qualifications component; a set of quality standards comprise each quality component. Quality **indicators** are metrics that can be measured or verified for each of the quality standards. A given quality standard could have one or multiple quality indicators that represent it in a QRIS. For example, in the category of staff qualifications, a standard may be “Teaching staff have specialized training in curriculum and assessment.” An indicator related to this standard may be “At least 50% of teaching staff have completed the two-course statewide curriculum training session on curriculum and assessment.”

**Table 1. Four Related Approaches to Validating a QRIS**

<table>
<thead>
<tr>
<th>Approach</th>
<th>Activities and Purpose</th>
<th>Typical Questions Approach Addresses</th>
<th>Issues and Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>Examine the validity of key underlying concepts</em></td>
<td>Assess whether basic QRIS quality components and standards are the “right” ones by examining levels of empirical and expert support.</td>
<td>Do the quality components capture the key elements of quality? Is there sufficient empirical and expert support for including each standard?</td>
<td>Different QRISs may use different decision rules about what standards to include in the system.</td>
</tr>
<tr>
<td>2. <em>Examine the measurement strategy and the psychometric properties of the measures used to assess quality</em></td>
<td>Examine whether the process used to document and verify each indicator is yielding accurate results. 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 psychometrically sound. Examine the relationships among the component measures to assess whether they are functioning as expected. Examine cut scores and combining rules to determine the most appropriate ways to combine measures of quality standards into summary ratings.</td>
<td>What is the reliability and accuracy of indicators assessed through program administrator self-report or by document review? What is the reliability and accuracy of indicators assessed through observation? Do quality measures perform as expected? (e.g., do subscales emerge as intended by the authors of the measures?) Do measures of similar standards relate more closely to each other than to other measures? Do measures relate to each other in ways consistent with theory? Do different cut scores produce better rating distributions (e.g., programs across all levels rather than programs at only one or two levels) or more meaningful distinctions among programs?</td>
<td>This validation activity is especially important given that some component measures were likely developed in low-stakes settings and have not been examined in the context of QRIS.</td>
</tr>
<tr>
<td>Approach</td>
<td>Activities and Purpose</td>
<td>Typical Questions Approach Addresses</td>
<td>Issues and Limitations</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------</td>
<td>--------------------------------------</td>
<td>------------------------</td>
</tr>
</tbody>
</table>
| **3. Assess the outputs of the rating process** | Examine variation and patterns of program-level ratings within and across program types to ensure that the ratings are functioning as intended.  
Examine relationship of program-level ratings to other quality indicators to determine if ratings are assessing quality in expected ways.  
Examine alternate cut points and rules to determine how well the ratings distinguish different levels of quality. | Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?  
Do rating distributions vary by program type, e.g., ratings of center-based programs compared to ratings of home-based programs? Are current cut scores and combining rules producing appropriate distributions across rating levels? | 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. |
| **4. Examine how ratings are associated with children’s outcomes.** | Examine the relationship between program-level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes. | Do children who attend higher-rated programs have greater gains in skills than children who attend lower-quality programs? | Appropriate demographic and program level control variables must be included in analyses to account for selection factors.  
Studies could be done on child and program samples to save resources.  
Findings do not permit attribution of causality about QRIS participation but inferences can be made about how quality influences children’s outcomes. |
Table 2. Data Needs, Data Sources and Analysis Methods for Selected Studies

<table>
<thead>
<tr>
<th>Approach</th>
<th>Data needed</th>
<th>Data sources</th>
<th>Analysis methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examine the validity of key underlying concepts</td>
<td>Evidence about the relationship between key quality standards and desired outcomes.</td>
<td>Empirical literature on how proposed components contribute to high quality care and improved child outcomes. Experts in early childhood education who can provide input on the quality standards and indicators.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expert opinions about proposed quality standards and indicators.</td>
<td></td>
<td>Synthesis of available data relating to each component; Analysis of degree to which evidence meets criteria for relatedness; Consensus process; Decision rules that specify the value of components without an established evidence base.</td>
</tr>
<tr>
<td>2. Examine the measurement strategies and psychometric properties of the measures used to assess quality</td>
<td>Rating data from participating programs.</td>
<td>Most such data are collected as part of program ratings.</td>
<td>Distribution of provider scores on a given component; Correlations among components; Correlations of selected components with other measures.</td>
</tr>
<tr>
<td></td>
<td>Data from additional quality measures.</td>
<td>Additional quality measures may be collected to allow comparisons with measures being used in the QRIS.</td>
<td></td>
</tr>
<tr>
<td>3. Assess the outputs of the rating process</td>
<td>Program-level ratings from participating programs.</td>
<td>Most of the necessary data are collected as part of program ratings.</td>
<td>Examination of rating distributions by program type; Correlations of program ratings with other measures; Changes in rating distributions using different cut scores.</td>
</tr>
<tr>
<td></td>
<td>Raw scores from measures of quality that are included in the rating.</td>
<td>Another measure of quality may be administered to allow comparisons with program ratings.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data from additional quality measures that are not included in the rating.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Relate ratings to expected child outcomes</td>
<td>Program rating data from participating programs.</td>
<td>Program rating data are collected as part of program ratings.</td>
<td>Estimate the relationship between program ratings and child outcomes.</td>
</tr>
<tr>
<td></td>
<td>Assessments of child functioning.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Approach 1: Examine the validity of key underlying concepts.** This approach involves examination of the elements or concepts that are to be included in program ratings. It is an important validation activity because it provides the foundation for the quality components, standards and indicators that together will produce program-level ratings and that will be the focus of quality improvement activities. Together, the components included in ratings, (e.g., staff qualifications, learning environment, family engagement) define quality for the QRIS. This validation activity provides justification and support for the elements of the QRIS. If the examination includes stakeholders, the process can also promote buy-in for the QRIS.
This validation approach asks whether quality components, standards and indicators included in a QRIS are the “right” ones, and is similar to what is proposed in the Toolkit, under Validating Quality Standards (Lugo-Gil et al., 2011). Because this effort addresses the cornerstone concepts and measures of the QRIS, it ideally would be conducted prior to the implementation of the QRIS.

For QRISs, the key concept is quality of care. The quality of care in early childhood education and school-aged care (ECE-SAC) programs is a complex, multi-dimensional construct; this complexity is amplified in centers by the fact that programs are comprised of multiple classrooms staffed by multiple individuals. Quality can be operationalized using a number of specific quality components. However, most QRISs have adopted similar ones. The QRIS Compendium found that six quality components were included in the majority of the 26 QRIS that were examined (Tout et al., 2010). These categories include licensing compliance (26 QRISs), classroom environment (24 QRISs), staff qualifications (26 QRISs), family partnership (24 QRISs), administration and management (23 QRISs) and accreditation (21 QRISs). Three categories—curriculum (14 QRISs), ratios and group size (13 QRISs), and child assessment (11 QRISs)—are included in half or just under half of the QRISs assessed. However, while similarities exist in the general quality components included in QRISs, the way in which each of these components of quality is measured varies substantially.

One activity that can help to validate a QRIS’ underlying concepts involves assessing the degree to which the quality components in the QRIS rating include standards and indicators that have an empirical base linking them to key program, family and child outcomes. This assessment might include an examination of the degree to which each element as operationalized in the QRIS is viewed by experts as a valid measure of the component. A number of states (including Delaware, Rhode Island, Minnesota and Virginia) have used a systematic expert review process to help identify which quality components (and the standards and indicators that comprise each component) to include in their QRIS. Attention might also be paid to the views of programs and parents about the degree to which selected components reflect their priorities. For example, focus groups with parents were conducted in Minnesota to inform the development of the final rating tool used in the QRIS pilot (Minnesota Department of Education and Minnesota Department of Human Services, 2007).

Another activity which is part of this approach involves examining the research literature to determine the level of empirical support for each proposed component. This review would examine the research base on the proposed standards and indicators selected to represent program quality. The review would weigh the existing evidence and provide arguments for why a particular quality component should be included or excluded from the QRIS.

Purdue University’s scientific review of the quality standards contained in Paths to Quality, Indiana’s QRIS, demonstrates this approach. The overall goal of the review was to conduct an “external evaluation of the scientific validity” of the Paths to Quality standards (Elicker et al., 2007). The study included review of available evidence for the importance of each of the four quality components—Health and Safety, Learning Environment, Planned Curriculum, and National Accreditation—and the relationship of the standards and indicators of each component to other measures of quality and to children’s development and well-being. The review used standards of evidence to classify each proposed indicator. For example, one or two well-designed studies that supported the indicator was classified as “some evidence;” “substantial evidence” required more than five such studies. For three-quarters of the indicators, researchers found “substantial evidence” that they supported children’s development.
Like many validation activities, such reviews ideally would be updated from time to time to determine if revisions to the QRIS would be advisable in light of new research findings. Such a review might utilize such tools as the *QRS Compendium* (Tout et al., 2010) or *Caring for Our Children* (AAP/APHA/NRC, 2011) as well as other recently published findings.

**Approach 2: Examine the measurement strategies and the psychometric properties of the measures used to assess quality.** A second type of validation effort focuses on the attributes of the individual measures in the QRIS as well as on the way in which the measures are combined to produce the summary rating of program quality. This approach is similar to what is discussed in the QRIS Evaluation Toolkit under *Validating the Construction of Quality Levels* (Lugo-Gil et al., 2011). This approach addresses how well the measures are working in the context of the QRIS. These efforts ask questions such as, “is there evidence that a given indicator measures what it purports to measure?” “If it claims to have a specific number of dimensions, do we find those dimensions in our data?” “Is there sufficient variance in scores on this indicator to justify its inclusion in the QRIS?” “Do scores on the indicator covary in expected ways with other measures of quality?”

Efforts to address these issues might involve an assessment of the distribution of participating provider scores on a given rating element. For example, in Zellman et al.’s (2008) evaluation of Colorado’s QRIS, initial work revealed that the measure of family engagement then in use produced very little variation across programs; all programs achieved the highest score possible on this measure. This meant that the QRIS was expending substantial resources to collect data on a measure that did not differentiate among programs. Another validation activity might involve an assessment of the relationship of a given indicator to other indicators of quality, both those included in the QRIS and others. In such studies, it is important to look at the degree of correlation found; ideally, measures would be moderately correlated so that each measure provides some non-redundant program quality information (see Zellman et al., 2008 for an example). Correlation patterns also should make sense. For example, two measures of interaction quality should be more closely related to each other than to a measure of ratios. If such studies reveal for example that the correlation between ratios and interaction processes is very high, this result might argue for eliminating one or the other indicator from the QRIS, as they may not be providing additional information (although some QRISs include certain elements to ensure that they are paid attention to, even if their psychometric properties are not ideal).

The research literature provides limited guidance concerning the most appropriate ways to combine measures of quality elements into summary ratings (Lugo-Gil et al., 2011; Tout et al., 2009; Zellman et al., 2008). Yet this process is crucial to producing meaningful program quality ratings, which are the key output of the rating process. States that are collecting and combining data could use these data to conduct studies that examine the effects of altering cut scores or combination rules, much as Karoly and Zellman (2012) have done in a “virtual pilot” for California’s QRIS, using data collected for another purpose, or as was done in studies in Minnesota (Tout et al., 2011) and Kentucky (Isner et al., 2012). These efforts will help QRIS designers and policy makers consider how well indicators are working, which indicators appear to be picking up variations in quality, and how closely different indicators relate to each other.

A number of other existing studies examine the properties of proposed QRIS indicators and can provide guidance to QRIS validation efforts (Scarr, Eisenberg, & Deater-Decker, 1994; Zellman & Perlman, 2008; Tout et al, 2011; McWayne & Melzi, 2011). Additionally, tools exist to help QRIS stakeholders review the options for QRIS measures and to support decision-making about the inclusion of new measures. For example, a Quality Measures Compendium is available and updated on a regular basis (Halle, Vick-Whittaker, & Anderson, 2010). If promising new measures are developed, it might be worthwhile to examine the performance of a new measure against the measure in current use.
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 that are assigned to providers who undergo a rating. Studies conducted under this approach examine the degree to which the quality levels in the QRIS are meaningfully distinct from each other. The results of these studies may indicate that measures, cut scores, or rules for combining measures need changing in order to distinguish quality levels effectively. Because these studies can result in proposals for significant changes to the composition of 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 program-level score that is the final output of the rating process. Studies conducted as part of this approach ask questions like, “are providers that received four stars actually providing higher quality care than those that earned three stars?” Studies using this approach may also address questions about cut scores, e.g., “do different cut scores produce dramatically different program-level ratings, and if so, which cut scores produce distributions that most closely relate to other measures of quality?” These studies typically rely on a measure of quality not included in the QRIS to make this assessment, and examine whether assessments on both measures vary in predictable ways.

The University of Southern Maine is conducting a validation study of Maine’s QRIS to assess similarities and differences across program ratings; the study is also examining what if any differences exist between similar types of programs at different step levels (see Lahti et al., forthcoming, for further details on this study and several others.) For example, researchers in Maine administer the Environment Rating Scales (ERS; Harms & Clifford; 1989; Harms, Clifford & Cryer, 2005; Harms, Cryer & Clifford, 2006; Harms, Cryer & Clifford, 2007 ), which are not used to establish a rating in Maine’s QRIS, and examine whether there are statistically significant differences in ERS scores between programs at different rating levels. These findings help program designers determine if the quality levels determined by QRIS ratings relate in expected ways to an external measure of global quality.

As a second example of validation studies using this approach, Karoly and Zellman (2012) used data collected for another purpose to model some of the features of a newly-designed California QRIS. The data come from a 2007 survey of center-based providers that is representative of the state. Observations were conducted in 251 centers serving children birth to 5. The purpose of this “virtual pilot” study was to determine the likely distribution of programs across QRIS tiers using specified cut points, examine the association among quality components, and to identify “outlier” quality elements on which otherwise well-rated programs tend to score poorly. This information is very valuable at the design phase; data on “outlier” elements is particularly helpful in understanding what it will take for programs to improve their rating in a QRIS that uses a block design to designate ratings (in which all indicators at one level must be met before a rating at the next level is possible). By examining such things as the relationship between scores on the Classroom Assessment Scoring System (CLASS; Pianta, La Paro & Hamre, 2008) and the Early Childhood Environment Rating Scale - Revised (ECERS-R; Harms, Clifford & Cryer, 2005), and the relationship between staff education and training and other measures of quality, the work can help policymakers assess the value of different measures of quality, provide input into establishing cut scores, and suggest targets for technical assistance efforts.
Other states also have conducted validation studies that focus closely on differences in QRIS levels. For example, Pennsylvania has studied programs participating in the Keystone STARS QRIS (Fiene, Greenberg, Bergsten, Fegley, Carl, & Gibbons, 2002; Barnard, Smith, Fiene, & Swanson, 2006; OCDEL (Office of Child Development and Early Learning), 2010; Manlove, Benson, Strickland, & Fiene, 2011) to determine if their program ratings were indicative of quality differentials across program types and services. Similarly, recent work in Indiana (Elicker, Langill, Ruprecht, Lewsader & Anderson, 2011) found that ERS scores varied with program-level ratings, while research in Minnesota found significantly higher scores on the ERS and CLASS only between the highest level (4-star) of the QRIS and the other rating levels (2- and 3-stars) (Tout et al., 2011). These findings are being used by program developers to make needed adjustments to quality indicators, metrics and cut scores.

**Approach 4: Relate ratings to children’s development.** A fourth approach to validation focuses on children’s development. It is similar to the Toolkit’s Linkages between quality levels and desired outcomes, although it focuses more narrowly on child outcomes. For QRISs, the logic model asserts that higher quality care will be associated with better child outcomes. Therefore, one important piece of validation evidence concerns whether children make greater developmental gains in programs with higher program-level ratings than in programs with lower ratings.

Studies using this approach do not attempt to identify causal linkages between **QRIS participation** and children’s outcomes. Instead, they examine whether the QRIS ratings and quality components that comprise the ratings are related in expected ways to measures of children’s development. Appropriate designs and controls could allow causal inferences to be made about how **quality** (as measured and rated by the QRIS) influences children's outcomes.

To date, few QRIS validation studies have incorporated children’s outcomes as they are costly and difficult to conduct. 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 vary between children whose families have selected different programs. Additional challenges include recruitment of programs and children across all quality levels; availability 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.

In Missouri, children who participated in programs with higher quality ratings showed significantly greater gains on measures of social-emotional development compared to children in programs with lower ratings (Thornburg et al., 2009). These effects were even more pronounced for low-income children. However, in an evaluation of Colorado’s QRIS, linkages between the ratings and children’s outcomes were not found (Zellman et al., 2008). Recent reports from Indiana (Elicker, Langill, Ruprecht, Lewsader, & Anderson, 2011) and Minnesota (Tout et al., 2011) found no consistent relationships between program ratings and measures of child outcomes. A number of possible explanations were offered for the lack of expected linkages, including overall low levels of quality in participating QRIS programs (perhaps not meeting a threshold of quality necessary to detect linkages with child outcomes; see Zaslow et al., 2010 for further discussion of quality thresholds) and a lack of variation among participating programs and families. Yet, even with these limitations, program administrators in both Indiana and Minnesota have used the findings to recommend changes to the structure and content of the QRIS.
Developing a Validation Plan

Given the complexity of validation, it is advisable to develop a plan for system validation as early as possible in the QRIS design process. Ideally, the validation plan will be part of a larger evaluation plan designed to address a wider range of important questions the answers to which will guide refinement of the QRIS and its implementation. The plan should include the key questions that will be addressed and the methods to be used to address each one. One advantage of developing a plan early is that it may highlight opportunities to conduct a number of the proposed efforts as part of the implementation of the QRIS itself or as part of planned evaluation activities. A comprehensive approach to validating a QRIS ideally will include studies under each of the four approaches described above. Table 3 outlines issues in the timing of validation studies, discusses their relative cost, and suggests strategies for addressing validation questions if resources do not permit the implementation of validation studies.

Table 3. Considerations in Developing a Validation Plan

<table>
<thead>
<tr>
<th>Approach</th>
<th>Timing and Duration</th>
<th>Cost considerations</th>
<th>Options to consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examine the validity of key underlying concepts</td>
<td>Ideally conducted prior to QRIS implementation.</td>
<td>Relatively inexpensive. This work can be contracted to a local university, consultant or research firm.</td>
<td>Many states are using similar concepts and measures; their efforts will provide useful information.</td>
</tr>
<tr>
<td></td>
<td>Study should be able to be completed within 3-6 months.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Examine the measurement strategies and psychometric properties of the measures used to assess quality</td>
<td>Must wait until ratings are implemented, although individual measures themselves might be available from other sources and could be examined earlier.</td>
<td>Depends on data quality and amount of analysis. Additional measures will increase costs, particularly if the measure is observational.</td>
<td>Can rely to some extent on existing research on each of the components. Consider using available data for a “virtual pilot.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Assess the outputs of the rating process</td>
<td>Must wait until ratings are implemented. Once data are available, several studies could be conducted using the same data set.</td>
<td>Depends on data quality and amount of analysis. Additional measures will increase costs, particularly if the measure is observational.</td>
<td>This work is state system-dependent so is not readily borrowed, though lessons learned about structure and cut-points can be shared across QRISs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Relate ratings to children’s development</td>
<td>Best to launch these studies when the QRIS rating process is stable and adequate numbers of programs have been rated.</td>
<td>Costs for the collection of child data are very high. Study could be done just with one cohort of children and two rounds of data collection (fall and spring) to assess developmental gains.</td>
<td>Requires significant funds, a powerful research design, and research expertise. Sampling children and programs will substantially reduce costs.</td>
</tr>
</tbody>
</table>

Summary and Conclusions

Validation is a complex, ongoing, iterative process. The objective of validation activities is to understand whether the rating process is able to distinguish among programs of different quality levels and whether program ratings are associated in meaningful ways to children’s outcomes.

Validation activities help to determine whether key design decisions are working well in practice. States and localities that have implemented QRISs are expending substantial resources to train raters, fund ratings, support various forms of technical assistance, and provide a range of improvement incentives. All of these efforts assume that the ratings are accurate and the system is performing as intended. QRIS design decisions often rely heavily on the judgments of experts and on colleagues in other states, because there is limited empirical data on which to base them. For this reason, it is critical for states to set in place a process for assessing how well the design decisions underlying the system are working. Validation activities do this.

Ideally, validation is an ongoing process based on a carefully designed validation plan. The plan should include all four validation approaches, although resource constraints may limit these efforts, and may particularly limit studies that include child outcomes. A good validation plan, thoughtfully developed and implemented, can provide information critical to improving the system at many points in the process, and increase the odds of its ultimate success. Validation is unquestionably challenging, but no more so than the launch and operation of a QRIS or its evaluation. The networks and references in the next section can help states develop a deeper understanding of validation approaches and help them construct and implement validation plans that address stakeholder and system needs and produce timely and valuable information.

Resources and References

Resources

INQUIRE – Quality Initiatives Research and Evaluation Consortium

The purpose of INQUIRE is to support high quality, policy-relevant research and evaluation on Quality Rating and Improvement Systems and other quality initiatives by providing a learning community and resources to support researchers and evaluators. INQUIRE also provides input and information to state administrators and other policymakers and practitioners on evaluation strategies, new research, interpretation of research results, and implications of research for practice. Research briefs are available on topics related to QRIS evaluation issues and strategies.

CCEERC – Child Care and Early Education Resource Connections
http://www.childcareresearch.org/ search under Quality Rating and Improvement Systems.

This site has many additional reports and resources, such as:


This resource list is an annotated bibliography of selected research focused on the design, implementation, and evaluation of Quality Rating Systems and Quality Rating and Improvement Systems in early childhood and after school settings.
The Child Care Quality Rating System (QRS) Assessment

Describing 26 Quality Rating Systems nationwide (19 statewide and 7 local or pilot), the Compendium presents comprehensive information through cross-QRS matrices and individual QRS profiles.


The QRS Assessment Toolkit will provide guidance, recommendations and evaluation support on a range of topics including: development of a logic model and research questions, evaluation design and methods, and selection of measures.

QRIS National Learning Network
http://qrisnetwork.org/

The Network provides information, learning opportunities, and direct technical assistance to states that have a QRIS or that are interested in developing one. Its National Resource Library assists states in learning more about QRIS and their elements and in QRIS planning. The library contains, toolkits, handouts and published documents on a variety of searchable topic areas.

The Networks’ State Resource Library contains detailed QRIS implementation information, including training guides, forms, and technical assistance materials that individual states have developed for their QRIS.

State QRIS Contacts who have agreed to serve as peer resources for one another are listed, as are Technical Assistance Providers.

Additional Resources

This report will provide case studies of four states that have undertaken validation studies in their respective states. This report provides validation and evaluation approaches, identification of similar QRIS standards amongst the four states, description of cross case analysis QRIS validity issues and the results of the validation conceptual model from this brief examining the following: concepts of quality, measures used to assess quality, outputs or scores of the rating process, and if ratings are related to expected outcomes. It is the companion document to supplement this guide in which four states validation experiences are highlighted.


The *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* was compiled by Child Trends for the Office of Planning, Research and Evaluation of the Administration for Children and Families, U.S. Department of Health and Human Services, to provide a consistent framework with which to review the existing measures of the quality of early care and education settings. The aim is to provide uniform information about quality measures. It is hoped that such information will be useful to researchers and practitioners, and help to inform the measurement of quality for policy-related purposes.

**References**


Endnotes

I Validity is not attached to a measure, but to a measure used for a particular purpose in a particular context. This means that measures which may be valid for one use must be validated again for use in a different context (AERA, APA, & NCME, 1999). Measures developed in low-stakes contexts, e.g., for use in research or program self-assessments, must be validated again in high-stakes contexts because those being assessed may react in high-stakes contexts in ways that could undermine the meaningfulness of interpretations derived from those measures (AERA, APA, & NCME, 1999).

II Some components such as parent involvement have been included in QRISs even when strong empirical support of the ability of measures to distinguish among programs of different quality was lacking because designers believed that if they were not, programs would ignore these components in favor of measured ones.

III Random assignment of children to programs with different quality ratings is not possible in QRIS. Alternative analytic approaches must be used that employ adequate controls for selection bias. See Zellman and Karoly (2012) for further discussion of this approach.

IV This column recognizes that state budgets are limited and validation is rarely seen as the highest priority. Ideally, states might combine data and efforts to conduct some of these studies.

V Ideally, states might combine data and efforts to conduct some of these studies.

VI However, as noted above, measures collected in low-stakes and high-stakes settings cannot be assumed to be comparable.

VII It may be possible to use existing data to test assumptions and measures. See, for example, Karoly and Zellman (2012), for a description of such work in California.
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

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 requirements 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 accomplish 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 support 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 1

**Pennsylvania Child Development Program Evaluation**

*Specific Items Within Identifed General Areas*

<table>
<thead>
<tr>
<th>General Requirements</th>
<th>4. Child abuse reporting procedures</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Staffing Standards</th>
<th>Staff requirements</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employee Records</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Evidence of qualifications and references for staff</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Building &amp; Site</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Appropriate indoor and outdoor</td>
<td>materials</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equipment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Condition and placement of equipment</td>
<td>2. Swimming regulations</td>
</tr>
<tr>
<td>3. Napping rules</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Program for Children</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Food &amp; Nutrition</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transportation</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child Health</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Staff Health</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Procedures for staff illness</td>
<td>2. Physical requirements for infant caregivers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedures &amp; Applications</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Child Records</th>
<th></th>
</tr>
</thead>
<tbody>
<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**

![Diagram](image)

**FIGURE 1**
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-

---

**The Indicator Checklist Approach**

![Diagram of the Indicator Checklist Approach](image)

**Figure 2**
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).
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**

**A. GROUP SIZE AND ADULT CHILD RATIOS:**
- INFANTS 1 STAFF TO 5 CHILDREN
  10 INFANTS IN A GROUP
- TODDLERS 1 STAFF TO 4 CHILDREN
  8 TODDLERS IN A GROUP
- PRESCHOOLERS 1 STAFF TO 10 CHILDREN
  20 PRESCHOOLERS IN A GROUP
- SCHOOL AGE 1 STAFF TO 15 CHILDREN
  30 SCHOOL AGE CHILDREN IN A GROUP

**B. SUFFICIENT SPACE—MINIMUM OF 40 SQ FT PER CHILD;**

**C. EQUIPMENT IS EASILY ACCESSIBLE TO CHILDREN;**

**D. ALL VEHICLES ARE EQUIPPED WITH AGE-APPROPRIATE SAFETY CARRIERS;**

**E. CLEANING MATERIALS ARE INACCESSIBLE TO CHILDREN;**

**F. EMERGENCY CONTACT INFORMATION IS AVAILABLE FOR ALL CHILDREN;**

**G. ALL STAFF HAVE HAD PERIODIC HEALTH APPRAISALS;**

**H. ACTIVITIES PROMOTE:**
- DEVELOPMENT OF SKILLS
- SELF-ESTEEM
- POSITIVE SELF-IDENTITY
- CHOICE OF ACTIVITIES.

(Fieme, 1984)

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 mathematical formula (phi coefficient) to select those items from the long questionnaire that are most useful in distinguishing be-
between 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

![Diagram showing the steps to construct an indicator checklist](FIGURE 3)
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 intrument 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.


A Systematic Approach to Child Care Regulatory Review, Policy Evaluation and Planning to Promote Health and Safety of Children in Child Care

A Manual for State and Local Child Care and Maternal and Child Health Agency Staff

Abbey Griffin, Ph.D.
Richard Fiene, Ph.D.
PROJECT HISTORY

ZERO TO THREE\National Center for Clinical Infant programs completed a three-year child care quality improvement project in 1994. The United States Maternal and Child Health Bureau (MCHB) offered four grants to organizations proposing different ways to use the national guidelines published by the American Public Health Association (APHA) and the American Academy of Pediatrics (AAP), Caring for Our Children: National Health and Safety Performance Standards: Guidelines for Out-of-Home Child Care Programs (APHA/AAP, 1992, funded by MCHB).

The APHA/AAP child care guidelines represent a major national effort to offer states a complete reference to current knowledge in the field of health, safety, and the policy and programmatic issues that assure protection for all children in the regulated child care system.***

Staff and expert consultants (see Attachment A for list and contact information) in consultation with state child care and maternal and child health administrators designed a state-level comprehensive regulatory review and planning process. The project addressed four components of a state’s child care quality assurance (or consumer protection) system: (a) child care regulations; (b) child care monitoring system; (c) child care training and public education; and (d) data collection and analysis for policy planning.

In 1991, ZERO TO THREE’s staff began working intensively with two states, Florida and Utah. Florida and Utah were similar in having a child care resource and referral system that collected data and, in addition, had substantial responsibilities for child care training and public information. They differed, however, in demographics, in the length and depth of state regulations, and in the balance of state versus county, or district, control of monitoring, training, and public education. Both states were preparing for a major regulatory review as required under the Child Care and Development Block Grant.

Each state’s first objective was to use the national guidelines to assess the adequacy of existing state child care regulations and refer state planners to the most current knowledge in the areas of health and safety. Second was to study the efficiency and effectiveness of the state’s child care monitoring systems. Developing monitoring tools to improve monitoring was the third objective. Fourth, the project invited state administrators

---

* Caring for Our Children was updated in 1994 and is now available from the National Center for Education in Maternal and Child Health (see Appendix A for address and technical assistance resources of the center).

** In this manual, we will refer to the APHA/AAP document as "the national standards."
to examine a number of options for improving quality in child care, including guidelines and training for child care monitoring staff, training for caregivers, and public awareness.

Experiences in Florida and Utah have offered rich insights into the benefits of the full model and its potential for informing the policy planning process. As with any pilot, the pitfalls and barriers are also found. For example, it takes time to initiate, complete, and then institutionalize a new set of procedures. It requires a shared commitment on the part of all the key players, communication of the purpose and rewards to those in the field who will participate, and negotiating agreements that will increase the efficiency of the process and the usefulness of its results.

Recognizing that institutionalizing a new process requires time, we developed this manual. This manual summarizes the goals, rationale, and procedures for each step of the process. Reflecting our experience in Florida and Utah, it is intended to offer guidance to those states in continuing the process as well as other states that want to develop a similar process.

ACKNOWLEDGMENTS

The Maternal and Child Health Bureau (MCHB) promoted pilot projects to demonstrate different ways that the national guidelines could be used effectively by states. The continuing work of the National Center for Education in Maternal and Child Health to build a national child care regulation data base, offer on-line technical assistance and reference searches, update and reprint the national guidelines, and conduct comparisons between a state's regulations and the national guidelines demonstrates MCHB's determination to make a difference for children and child care through these guidelines.

ZERO TO THREE is grateful to MCHB, and in particular Dr. David Heppel, Dr. Phyllis Stubbs and Denise Sofka, who realized the need and the timeliness of creating national consensus on guidelines to promote the health, safety and development of children enrolled in child care. It was an MCHB grant that supported the APHA's and AAP's three year effort to develop the national guidelines.

This manual is dedicated to those Florida and Utah state administrators who coordinated the work, and to those local administrators and child care resource and referral agency staff, program administrators, parents and others who helped complete the various phases of the project. It is our hope that the manual will help them institutionalize and improve the model in the years to come.

The project was guided by national leaders in the fields of early care and education and health, Dr. Susan Aronson, Dr. Richard Fiene, Gwen Morgan and Pauline Koch. Their wisdom and deep commitment to early care and education were invaluable. Finally, ZERO TO THREE staff members Helen Keith and Monique Amos were vital to the success of the
project. Helen Keith was Co-Director of the project and, as a former state administrator in Vermont, provided insights into state systems and an empathy for the complexities of state administration that kept both the state and national teams going through the more difficult stages of the process. Monique Amos contributed research, organization, planning, coordination, and magic on the computer.
A SYSTEMATIC APPROACH TO CHILD CARE REGULATORY REVIEW, POLICY EVALUATION AND PLANNING TO PROMOTE HEALTH AND SAFETY OF CHILDREN IN CHILD CARE: A MANUAL FOR STATE AND LOCAL CHILD CARE AND MATERNAL AND CHILD HEALTH AGENCY STAFF

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section I</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>To present the need for a systematic approach to child care policy evaluation, planning and quality improvement; and describe the purpose and content of the manual.</td>
<td></td>
</tr>
<tr>
<td>Section II</td>
<td>Regulatory Analysis</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>To describe ways state administrators can use the national health and safety guidelines; and outline strategies for comparing the state's child care regulations to the national guidelines.</td>
<td></td>
</tr>
<tr>
<td>Section III</td>
<td>State Regulatory Compliance Study</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>To present an approach to data collection by state administrators to assess child care program compliance with state regulations, identify systemic problems and specific program or geographic areas in need of assistance, and make recommendations for allocating limited resources.</td>
<td></td>
</tr>
<tr>
<td>Section IV</td>
<td>Field-Tested Weighted Indicator Monitoring Tool</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>To describe and present the steps required to develop a statistically reliable weighted indicator checklist for monitoring compliance in child care facilities that will improve the efficiency and effectiveness of the state's child care monitoring and data collection procedures.</td>
<td></td>
</tr>
<tr>
<td>Section V</td>
<td>The Data Display Chart and The Recommended Action Planning Chart</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>To present two chart formats that will assist state administrators by recording all collected data, decisions made over the course of the review process and final recommendations.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endnotes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>References</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appendix A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resource List</td>
<td></td>
</tr>
</tbody>
</table>
SECTION I

Introduction to the Manual
SECTION I
INTRODUCTION TO THE MANUAL

THE NEED FOR A SYSTEMATIC APPROACH TO STATE CHILD CARE REGULATION, MONITORING, DATA COLLECTION

State child care regulations represent a consensus among citizens and their state and local government representatives about what is to be considered essential to protecting the health, safety and well-being of children in regulated child care settings. The criterion for promulgating regulations, codes and special requirements is most often "protection of children from harm." Thus, state and local governments have an obligation to the public to assure child care program compliance with legal requirements to protect children from harm. States need data from their monitoring system to insure that their obligations to families are met.

Scant data are available to demonstrate either the adequacy of state and local child care rules or compliance with those rules across a state. What exists highlights the following weaknesses:

- Changes in child care regulations are too often the result of crises or pressure from some segment of the population rather than a systematic process that includes an assessment of program compliance data or a review of expert knowledge.¹

- Child care regulations are written and formatted to meet administrative and legislative specifications. They tend to be difficult to read and understand, a circumstance which leads to misinterpretation and inconsistent application by both monitors and child care providers.²

- Child care licensing staff cannot adequately monitor programs for compliance for a combination of reasons: (a) large caseloads; (b) lack of training and guidance materials; and (c) lengthy review procedures.³

- Child care rule making, enforcement and data collection processes are complicated by the number of state and local agencies that may have authority to promulgate rules and, in some cases, monitor for compliance, as well as by inconsistent interpretation of regulations.⁴

Changes in child care regulations are too often the result of pressure on state or local legislators, media attention to a dramatic event in a regulated child care setting, fiscal crises, or changes in Executive Branch priorities. In one state, proposals by child care professional and advocacy organizations as well as state and local administrators to extend basic health and safety regulations, and building and fire codes to all forms of child care had languished for years, but when a two-year-old drowned at a license-exempt drop-in child care center at a shopping mall, the issue of unregulated or "exempt" categories of child care was quickly moved onto the legislative agenda.
There are many reasons why such issues are not fully debated until there is a crisis. Among them is the absence of systematic efforts to collect expert knowledge and state data to build a convincing case for extending a basic floor of protection to all children in group child care settings.* While the model presented in this manual is limited to the regulated system, it does set in place the kind of review process and data collection process states need in order to expand data collection to exempt and unlicensed/or unmonitored child care settings.

Research on child care has been consistent, to a degree rarely found across both large and small sample studies, in identifying child care quality indicators that promote healthy, safe environments for children. These indicators include: small group size, high staff-to-child ratios, staff education and specialized training, family involvement, health promotion for staff and children, safe physical environments both inside and outside, and continuity of caregiving, where relationships can develop over time among staff, family members and children.

There has not been, to date, a systematic process for reviewing and revising state child care regulations and requirements. Furthermore, as the Inspector General’s Report (Kusserow, 1990) stated, no national studies have examined state regulatory enforcement.5 The Inspector General’s Report outlines the following findings:

- States do not regulate child care in all types of settings.
- Patterns of state regulation are constantly changing.
- Frequent visits are the best way to ensure compliance. However, only a portion of facilities are monitored regularly.
- Legal sanctions are time-consuming and difficult to enforce, even in cases where children are at risk.
- Parental involvement is considered the first line of defense but may be the weakest link in the chain. (Kusserow, 1990 Part 1, p.ii)

The Inspector General’s Report cites several innovative approaches developed in Delaware, Minnesota, Pennsylvania, Washington and Wisconsin. They include the use of:

- new monitoring techniques (indicator checklists, interviews, field surveys and focus groups, inspector training and guidance materials);
- monetary incentives and penalties (financial incentives and fines);

---

* Group child care is defined here as more than four children care facility by someone other than the parents or relatives of the children.
• parent involvement (rules regarding parental involvement, consumer education materials, inclusion of parents and providers in regulatory review and enforcement policy decisions);

• training and technical assistance to child care monitors and practitioners (increasing the availability, accessibility and affordability of child care training, identifying needs and targeting training resources); and,

• regulatory guidance materials, consumer and public education (maximizing the use of state and local agency resources to provide health, safety and child development information, and materials to explain child care rules and provide examples of compliance).

This manual addresses three of the approaches. While legal sanctions and procedures are not covered, the planning process targets both regulatory (i.e., regulation and enforcement) and non-regulatory (i.e., training, technical assistance, and public education) approaches.

BENEFITS OF A SYSTEMATIC REVIEW AND PLANNING PROCESS

A routine and well-informed review of child care regulations, program compliance, and monitoring effectiveness offers important benefits to the state. States have limited funds to meet concurrent demands for increasing the supply of child care services, expanding the number of child care sites that can enroll children with disabilities, and improving the quality of child care available within their boundaries.

The approach presented in this manual offers state administrators the following benefits:

1. Bringing together state agencies that promulgate child care rules:
   - reduces duplication of effort;
   - increases clarity of child care rules;
   - simplifies the licensing process for child care providers;
   - encourages the sharing of informational, training and technical assistance resources among state agencies.

   - informs child care policy planning with current knowledge on health and safety in child care;
• highlights areas in which child care regulations fail to protect young children or promote their health, safety and development.

3. Completing a compliance study for all regulated child care at regular intervals (e.g., every three years):

• allows state child care administrators to strategically allocate limited human and financial resources to areas of the state and/or types of child care settings where compliance is low and/or resources for improving quality are weak;

• guides decision-making and long range planning by measuring the effects of both regulatory and non-regulatory quality improvement strategies (e.g., new rules, more frequent monitoring, training, technical assistance, public education).

4. Developing a statistically reliable short form for monitoring (i.e., a weighted-indicator checklist):

• decreases the amount of time monitors spend reviewing child care programs with a history of compliance;

• allows monitors to focus their efforts on child care programs that need technical assistance or corrective action.

5. Providing reliable data to state and local policy planners, legislators, and others increases efficiency in completing state and federal reports and program plans.

All states have a variety of settings that offer child care, including care by a relative or neighbor, care in small groups in a family child care home or in larger groups in group family child care homes, part-day programs, and child care centers. While all states address child care centers in their state regulations, some states exempt certain types of centers -- for example, those located in schools, on military bases, or in churches.

In this manual, we discuss a regulatory analysis model, which collects and analyzes data on only the children served in regulated, certified or registered, and monitored settings. However, it is hoped that once state and local agencies have institutionalized this model, data can be collected on license-exempt child care programs as well. The challenge of reaching those family child care programs and less formal group care settings that are not licensed or registered is more difficult to meet. Some states are already using census data and parent surveys as well as tracking parent payments using state child care vouchers to estimate the number of unregulated child care facilities and the ages of children served by them.
USING THE MANUAL

This manual describes the key elements and implementation of a systematic regulatory analysis, policy evaluation, and planning process to improve the content, efficiency, and effectiveness of state child care administration. Through several relatively easy data collection procedures and a review team process, state administrators can identify weaknesses in their child care quality assurance system which could, if unaddressed, put children at risk.

The purpose of the process is formative -- to strengthen and guide state and local child care policy decisions. The goal is to collect information needed for problem-identification and problem-solving that will lead to a more efficient, effective and cost-effective state quality assurance system.

The review process brings together administrators with authority to promulgate rules and administer the child care regulatory and monitoring and subsidy systems and other programs that have the resources and expertise to improve the health, safety and development of children in child care. The decision-making process considers the data, researches the cause/s of weaknesses seen in the data, as well as examining the existing and potential resources of the agencies involved in order to recommend the most effective and cost-efficient actions to take, including; rule changes, guidance materials that clarify the rules, training, and/or consumer education.

If the proposed regulatory review and policy evaluation and planning process is completed at regular intervals, state administrators can measure whether existing policies and administrative initiatives have been effective in improving child care program compliance. State and/or local administrators can also test the relationships between specific policies (e.g., training) and improved compliance. The data can greatly facilitate the preparation of Federal and state reports and be used to inform both the legislature and the public.

Section headings represent the major components of the process:

1. **Crosswalk/Comparison**, between state regulations and the national guidelines, used to identify gaps and weakness in the state’s existing child care regulations and provide a reference to expert knowledge and model language and policies.

2. **State Compliance Profile**, based on a random, demographically representative sample of recent local monitoring reports. Consists of both aggregate and specific analyses of compliance in regulated child care across the state and demonstrates relationships between variables of interest to the state (e.g., training) and compliance.

3. **Field-tested weighted indicator checklist**, based on both compliance data and a survey of representatives of the child care field and families using child care. Provides data needed to derive a statistically reliable weighted indicator checklist monitoring tool.
Weighted Indicator Checklist, a short monitoring instrument that is statistically reliable in predicting compliance and used to streamline the monitoring process.

4. Data/Action Planning Chart documents, data collected, team planning and decision-making at each review team meeting and facilitates the review and evaluation of the regulatory system as well as other state investments in the child care system (e.g., training, guidance materials, consumer education materials).

5. Recommendations Planning Chart, a tool for decision making with columns: (1) issue; (2) recommendations ("best fit" national guideline typed); (3) code for the national guideline; and, (4) action-to-be-taken columns.

The manual is designed to present a comprehensive process while making it easy to pull out individual elements. Each section is written as an independent unit having three parts:

1. An overview of the purpose, function and outcomes of that step;

2. A sample of forms, research instruments and reporting formats (with data); and

3. A technical description of design and statistical procedures with examples of state data presentations.

There are currently a number of resources available to assist state administrators in accomplishing the tasks outlined. A resource list is provided at the end of the manual (Appendix A).

THE AUDIENCE

The manual is written for state and local administrators and statewide child care and child health organizations who are looking for ways to improve the efficiency and effectiveness of their child care regulatory review, rule-making process and enforcement system.

The author strongly recommends that administrators of child care and maternal and child health policy divisions be included in the review process and long-term planning, which should consider training, guidance materials and consumer education as approaches to quality improvement.

The chart on the next page provides an overview of the process. The "inputs" include the people, information, systems and organizations that are involved in each component of the process. "Processes" identify each component and major contribution to the process. The last column, "Outcomes," identifies the benefits to the state of including each component.
SECTION II

Regulatory Analysis
SECTION II

REGULATORY ANALYSIS

PART 1: HOW DOES YOUR STATE MEASURE UP IN PROTECTING YOUNG CHILDREN IN CHILD CARE SETTINGS?


State child care regulations are often weak, unclear or absent in the following areas:

- Group size
- Adult-to-child ratios
- Acoustics
- Playground safety
- Parent access at any time of the day and parent involvement
- Complete and up-to-date child health and immunization records
- Complete and up-to-date caregiver health and immunization records
- Explicit reporting requirements for infectious disease and injuries requiring medical attention
- Provision and documentation of caregiver orientation and in-service training
- Explicit requirements for sanitary procedures, particularly when children are in diapers or learning to use the toilet but lack full bowel control
- Explicit requirements for the storage of toxic materials.

The North Carolina Department of Health conducted an audit of health and safety standards in 27 child care centers. In four centers, with a capacity of 58 children, there were no violations; but in 23, with a capacity of 1,119 children, 214 violations were found. Among the violations were playground and indoor area hazards, toxic chemicals accessible to children, unsanitary conditions, and incomplete child and staff records.6

Most injuries in child care are due to inadequate adult supervision, which is directly related to both staff-child ratios and group sizes. Toddlers have the highest incidence of injuries; yet in 45 out of 50 states, group size and staff-child ratios for this group (children 12 to 36 months of age) are the most out of congruence with national standards (i.e., APHA\AAP national health and safety guidelines and the accreditation standards set by the National Association for the Education of Young Children).

Over half of all injuries requiring medical attention occur in outdoor play areas. The most frequent and severe injuries are associated with falls to hard surfaces, with entrapment being the second most frequent cause. These risks can be prevented by assuring that all
child care facilities use adequate shock-absorbent materials under climbing equipment, sufficient in diameter to cover the fall zone, and understand what size equipment can safely be used by children at different ages and that close supervision is critical. Yet little guidance is given to families and child care providers about playground safety. In the majority of states, child care regulations specify only that outdoor equipment be "free of hazards."
PART 2: COMPARISONS BETWEEN YOUR STATE'S CHILD CARE REGULATIONS AND THE NATIONAL GUIDELINES

GOAL: TO ASSESS THE ADEQUACY OF STATE CHILD CARE REGULATIONS IN MEETING NATIONAL CHILD CARE HEALTH AND SAFETY GUIDELINES

NATIONAL GUIDELINES: (NATIONAL HEALTH AND SAFETY PERFORMANCE STANDARDS: GUIDELINES FOR OUT-OF-HOME CHILD CARE PROGRAMS) (APHA/AAP, 1992, updated and available through National Center for Education in Maternal and Child Health, Arlington, VA under this title)

These 980-plus model standards represent the consensus of a multidisciplinary panel of experts on what is needed to assure the safety, health, mental health and development of young children enrolled in child care centers and family child care homes. The national guidelines are formatted in 3 columns: guideline, rationale, and comment. The appendices offer model policies and descriptive materials.

Two Ways To Format The Comparisons Between Your State Child Care Regulations And The National Standards:

1. Comparison #1: State rules vs. national guidelines

What to do:
List all state regulations on the left and in a facing column list all individual national guidelines dealing with the same topic. Pick the key word/s to represent your state rule.

How to use:
This comparison clusters all national guidelines that are most closely related to the content and purpose of the state regulation. The format makes it easy to reference expert knowledge and language for creating new rules, updating, clarifying and/or strengthening existing rules.

2. Comparison #2: National guidelines vs. state rules

What to do:
List all national guidelines on the right and, in the facing column, indicate whether the state covers the topic.

How to use:
This format highlights topics that are covered, covered poorly or not covered in your state child care regulations. Using the citations to the national guidelines,
administrators can review topics that are poorly or not covered that they may want to consider for future rule-making. The topics and specific citations to the national guidelines can also be used to assess the content of available training or be a resource in the development of regulatory guidance or consumer education materials.
PART 3: TECHNICAL GUIDE FOR CONDUCTING COMPARISONS

The National Resource Center for Child Care Health and Safety can do the comparison for you if they have your most current child care regulations, rules, and codes entered in their computer database. If not, they can run a search. Call 703-524-7802. For more information about the National Resource Center or for referral information to Dr. Richard Fiene (see Appendix A)

PREPARING AND PRESENTING COMPARISON BETWEEN STATE CHILD CARE REGULATIONS AND THE NATIONAL GUIDELINES

1. List each of your state child care regulations (& codes or rules) in a column on the left. Identify each by reference code number, by section heading and by topic, leaving plenty of blank space between each.

2. Pick key words for the issue/s covered by each state regulation and use to search the Index of the national guidelines for all citations. Cluster the national guideline code letters and numbers by section in the right column.

3. The format used on the Sample Comparison #1 on the next page makes all national guidelines that relate to the topic addressed in the state child care rules easy to identify at a glance.
## SAMPLE COMPARISON #1
### CHILD CARE STANDARDS

### SAMPLE STATE

<table>
<thead>
<tr>
<th>I. Definition:</th>
<th>APHA/AAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
</tbody>
</table>

### II. Program Administration:

<table>
<thead>
<tr>
<th>A. Policy and Procedure</th>
<th>APHA/AAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>II, A, 1, a-f</td>
<td>AD 1 and AD 5</td>
</tr>
<tr>
<td>II, A, 2</td>
<td>AD 1 and AD 7</td>
</tr>
<tr>
<td>II, A, 3</td>
<td>AD 1</td>
</tr>
<tr>
<td>II, A, 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Management</th>
<th>APHA/AAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>II, B, 1</td>
<td>AD 3</td>
</tr>
<tr>
<td>II, B, 2</td>
<td>APP 33</td>
</tr>
<tr>
<td>II, B, 3</td>
<td>AD 4 and APP 33</td>
</tr>
<tr>
<td>II, B, 4</td>
<td>APP 6 and C-2</td>
</tr>
<tr>
<td>II, B, 5</td>
<td>APP 6 and C-3-h</td>
</tr>
<tr>
<td>II, B, 6</td>
<td>AD 34</td>
</tr>
<tr>
<td>II, B, 7</td>
<td>AD 39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Communication</th>
<th>APHA/AAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>II, C, 1</td>
<td>AD70</td>
</tr>
<tr>
<td>II, C, 2</td>
<td></td>
</tr>
<tr>
<td>II, C, 3</td>
<td>FA109</td>
</tr>
<tr>
<td>II, C, 4</td>
<td></td>
</tr>
<tr>
<td>II, C, 5</td>
<td>HP 94</td>
</tr>
<tr>
<td>II, C, 5, a&amp;b</td>
<td></td>
</tr>
<tr>
<td>II, C, 6</td>
<td>HP101</td>
</tr>
</tbody>
</table>

### III. Emergency Plans:

<table>
<thead>
<tr>
<th>III, A</th>
<th>FA 116, (Says 2 kits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>III, B</td>
<td>AD 34</td>
</tr>
<tr>
<td>III, C</td>
<td>APP 28</td>
</tr>
<tr>
<td>III, C, 1</td>
<td>APP28E</td>
</tr>
<tr>
<td>III, C, 2</td>
<td></td>
</tr>
<tr>
<td>III, C, 3</td>
<td>AD 31</td>
</tr>
<tr>
<td>III, C, 4</td>
<td></td>
</tr>
<tr>
<td>III, C, 5</td>
<td>AD 33</td>
</tr>
</tbody>
</table>
PREPARING AND PRESENTING COMPARISON BETWEEN THE NATIONAL GUIDELINES AND STATE CHILD CARE REGULATIONS

1. List each national guideline by section heading, reference code and topic in the left hand column.

2. Since states have a much smaller set of regulations to match, order your regulations into section headings comparable to those used by the national guidelines. Search your state child care regulations, identifying those that reflect the purpose and content of the national guideline. If such a rule is not present, indicate its absence in the right-hand column; if present, cite the state rule.

3. The purpose of this comparison is to identify gaps and major areas of regulatory concern that are poorly addressed in your state regulations. It is important to be circumspect when judging the congruence between your state regulation and the national guideline.

4. The format in the Sample Comparison #2 on the next page makes it easy not only to see where gaps in state regulations exist but also to identify those national guidelines the state team may want to consider.
SAMPLE COMPARISON #2
APHA/AAP NATIONAL HEALTH & SAFETY PERFORMANCE STANDARDS:
GUIDELINES FOR OUT-OF-HOME CHILD CARE PROGRAMS

<table>
<thead>
<tr>
<th>APHA/AAP</th>
<th>SAMPLE STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Staffing</strong></td>
<td></td>
</tr>
<tr>
<td><strong>1.1 Child: Staff Ratio and Group Size</strong></td>
<td></td>
</tr>
<tr>
<td>ST1</td>
<td></td>
</tr>
<tr>
<td>ST2</td>
<td>VI, A, 3</td>
</tr>
<tr>
<td>ST3</td>
<td>VI, I, 1, d/VI, J, 5 &amp; 6</td>
</tr>
<tr>
<td>ST4</td>
<td></td>
</tr>
<tr>
<td><strong>1.2 Licensure/Certification of Qualified Individuals</strong></td>
<td></td>
</tr>
<tr>
<td>ST5</td>
<td></td>
</tr>
<tr>
<td><strong>1.3 Qualifications</strong></td>
<td></td>
</tr>
<tr>
<td>ST6</td>
<td>V, B, 1</td>
</tr>
<tr>
<td>ST7</td>
<td></td>
</tr>
<tr>
<td>ST8</td>
<td></td>
</tr>
<tr>
<td>ST9</td>
<td></td>
</tr>
<tr>
<td>ST10</td>
<td></td>
</tr>
<tr>
<td>ST11</td>
<td></td>
</tr>
<tr>
<td>ST12</td>
<td></td>
</tr>
<tr>
<td>ST13</td>
<td></td>
</tr>
<tr>
<td>ST14</td>
<td></td>
</tr>
<tr>
<td>ST15</td>
<td></td>
</tr>
<tr>
<td>ST16</td>
<td></td>
</tr>
<tr>
<td>ST17</td>
<td></td>
</tr>
<tr>
<td>ST18</td>
<td></td>
</tr>
<tr>
<td>ST19</td>
<td></td>
</tr>
<tr>
<td>ST20</td>
<td></td>
</tr>
<tr>
<td>ST21</td>
<td></td>
</tr>
<tr>
<td>ST22</td>
<td></td>
</tr>
<tr>
<td>ST23</td>
<td></td>
</tr>
<tr>
<td>ST24</td>
<td>V, A, 3</td>
</tr>
<tr>
<td>ST25</td>
<td>V, A, 3</td>
</tr>
<tr>
<td>ST26</td>
<td></td>
</tr>
<tr>
<td>ST27</td>
<td></td>
</tr>
<tr>
<td>ST28</td>
<td></td>
</tr>
</tbody>
</table>
A SAMPLE WORKBOOK DISPLAY FORMAT

The following is a format for presenting state regulations or topics not covered in the regulations that the review team has decided (from the comparisons to the national standards) should be discussed. Preparing the workbook takes time but can be easily handled by clerical staff since all the relevant citations to both state rules and national guidelines are in the comparison studies. The workbook format does increase the efficiency of the review team since all relevant information is provided and the outline provides a clear process for discussion and decision-making.

-----------

SAMPLE WORKBOOK

State Regulation:
I. Program Requirements:
   C. Communication
      1. Incidents, such as injuries and medical conditions that occur while a child is in care, are reported to the parents on the same day they occur.

Related National Guidelines:
AD70: When an injury occurs in the facility that requires first aid or medical attention for a child or adult, the facility shall complete a report form that provides the following information:
   a) Name, sex, and age of the injured.
   b) Date and time of injury.
   c) Location where injury took place.
   d) A description of how the injury occurred.
   e) Part of the body involved.
   f) Description of any consumer product involved.
   g) Name of staff member responsible for the care of the injured person at the time of the injury.
   h) Actions taken on behalf of the injured following the injury.
   i) Name of person who completed the report.
   j) Name and address of the facility.

The injury report form shall be completed in triplicate. One copy shall be given to the child’s parent or legal guardian (or to the injured adult). The second copy shall be kept in the child’s (or adult’s) folder at the facility, and the third copy shall be kept in a chronologically filed injury log. This last copy shall be kept in the facility for the period required by the state’s statute of limitations.

AD69: For illness with onset while a child is attending (or a staff member is working in) a facility that potentially require exclusion (see Inclusion/Exclusion/Dismissal, on p. 80), the facility shall record the date and time of the illness, the person affected, a description of the symptoms, the response of the staff to these symptoms, who was notified (e.g., parent, legal guardian, nurse, physician), and their response.
AD72: The facility shall document that a child's parent or legal guardian was notified immediately of an injury or illness that required professional medical attention.

AD74: The required form shall be sent to the licensing agency or health department within 5 working days in the event of the death of a child, hospitalization of a child for an injury that occurred in child care, or hospitalization for a reportable communicable disease.

Gaps?

Recommended action/s (rule change, guidance materials, training, consumer education):
SECTION III

State Regulatory Compliance Study
SECTION III

THE REGULATORY COMPLIANCE STUDY

PART 1: CAN YOUR STATE COLLECT DATA THROUGH ITS MONITORING SYSTEM TO ASSESS COMPLIANCE IN CHILD CARE ACROSS THE STATE AND EXAMINE SPECIFIC POLICIES?

Many states do set standards above the floor of "harm to the child," using one or more of the following approaches. States can allocate resources to improve quality through training. They can provide technical assistance and disseminate information to families and child care providers through child care information and referral agencies or other community organizations. Many states pay higher child care rates to programs that meet standards of best practice set by the early care and education field (e.g., Criteria for Accreditation by the National Academy of Early Childhood Programs). These states need data to support these investments in better quality child care.

The regulatory compliance study estimates the degree to which child care programs meet state regulations. It provides state policymakers a profile of child care within the state's boundaries. It targets geographic and demographic differences, identifies specific rules with low compliance, and can alert policymakers to problems in specific types of child care settings. A specific policy -- for example, increasing training of child care staff - can be analyzed in order to verify whether that policy has, in fact, improved overall program compliance.

When the regulatory compliance study is done at regular intervals (e.g., every three years), policymakers can measure progress over time. This offers policymakers a clear history of what works, and what is still a problem.
PART 2: THE REGULATORY COMPLIANCE STUDY AND STATE COMPLIANCE PROFILE

GOAL: TO COLLECT, ASSESS, AND INCREASE THE EFFECTIVENESS OF STATE CHILD CARE REGULATIONS AND POLICIES BASED ON COMPLIANCE DATA COLLECTED THROUGH THE STATE/LOCAL CHILD CARE MONITORING SYSTEM

METHODOLOGY FOR COLLECTING DATA AND ASSESSING COMPLIANCE

1. Selecting a Sample of Recent Monitoring Records

The state licensing office collects a random sample of comprehensive monitoring records (i.e., not short forms, if used in the state for interim monitoring visits) from all licensing districts/regional offices, representing the full distribution of child care facilities across the state. A 10% randomly selected sample from each office is usually sufficient to represent the state.

A few questions to ask in planning data collection:

1) If policymakers have reason to believe that this sampling process will not achieve an accurate demographic picture of the state, specific instructions need to be given to the local monitoring staff. For example, if low-income children in urban areas are more likely to be in large child care centers, while in rural areas they are in small centers, then a larger sample from those urban districts can be collected.

2) If there is more than one local agency that monitors child care programs (e.g., child care monitors and public health nurses or environmental health inspectors), it is optimal to collaborate with these agencies at both state and local levels to collect a random sample of records from each. The extra effort required has several benefits:

- identifying duplication of rules simplifies the process, making monitoring and data analysis more effective;
- creating consistency in rule definitions, language used, and monitoring procedures makes it easier for child care providers to understand and comply with rules; and
- reducing duplication and increasing consistency is more cost-effective, making better use of each agency’s expertise and resources.

3) Finally, if monitoring records include anecdotal records and observations, a comprehensive checklist with all state regulations included will be needed. More reliable data will be collected if the study time line is extended to allow district or regional offices time to adapt to and complete a full monitoring cycle.
2. Coding Records

Develop a standard coding scheme to categorize records for variables needed to capture an accurate picture of the child care system in the state. For example, the state would ask licensing offices to code on each record variables the review team will want to analyze.

A sample list of variables to code:

1) district/region
2) characteristics of the geographic area (e.g., rural/urban)
3) characteristics of the population served (e.g., family income, racial or ethnic make-up)
4) characteristics of the child care program (e.g., not-for-profit/for profit; size of the program; percentage of subsidized children served)

Coding specific variables of interest related to state policy decisions:

1) level of training attained by the child care program administrator
2) staff compliance with training requirements (even if not in statute, licensing offices can get the information from child care programs and, increasingly, child care information and referral programs collect training data).

3. Review Process -- The Review Team

(Statistical procedures and sample data display formats are covered in Part 3 of this section: Technical Guide for Analyzing Compliance Data)

The review team should include representatives from all child care agencies (policy, licensing, and subsidy management) and other agencies that promulgate rules for child care. In order to avoid conflict, it is important that all participants agree that the data is for internal review, problem identification, policy evaluation and planning. Data from the compliance study should only be released when its accuracy has been established and all participants agree to its use.

Thoughts on selecting a review team:

1) Participants bring to the review process current knowledge from their field as well as data collected through their agencies.
2) Their different perspectives add insight into why compliance might be low on specific regulations.
3) Differences in regulatory language or monitoring procedures can be negotiated, increasing the effectiveness of the regulatory system.
4) Sharing resources to solve problems and increase child care compliance can be achieved by using or adapting existing educational and consumer education
materials, expanding dissemination capacities, and expanding the use of training materials and facilities.

4. The Review Process -- Phase One

The first phase of the review process examines the percentage of non-compliance on each regulation. The data are analyzed to highlight non-compliance in order to make it easy to identify problems.

The state review team must agree on a percentage of non-compliance above which action must be taken to improve compliance. For example, in one state with a small set of state regulations (165) any rule with a 10% rate of non-compliance was considered an issue; while in a second state, with 280 individual rules, 15% was considered too high.

Questions to answer regarding unacceptably low compliance scores on specific regulations:

1) Is the regulation unclear?
2) Are there differences in the way rules are interpreted or in monitoring procedures?
3) Can the problem be the result of a lack of state, local or program resources?
4) Is the problem greater in specific districts, in specific types of programs, in specific geographic or demographic areas?

In some instances, data from other sources may contradict study findings. For example, study data on child immunization rates might differ from public health immunization statistics. In that case, the problem might be one of child care program record-keeping, health department sampling, or improper monitoring.

5. The Review Process -- Phase Two

The second phase involves the analysis of specific comparisons to assess state policies.

Sample variables and questions to be asked are:

1) Do programs that meet all staff training requirements have a higher rate of compliance?
   (Note: comparisons are constructed to measure compliance on a specific rule, or set of related rules, against compliance on all regulations)
2) When the child care administrator has more education, is program compliance higher?
3) Are small, medium or large programs more likely to have a higher rate of compliance?
6. The Review Process -- Final Phase

The final phase requires discussion, problem-solving and planning. Both regulatory and nonregulatory strategies should be considered.

**Strategies to address identified problems fall into four categories:**

1) rule change/new rule
2) guidance material to clarify rule
3) training of child care and/or monitoring staff
4) consumer and public education material
PART 3: TECHNICAL GUIDE TO DATA COLLECTION AND STATISTICAL ANALYSIS

DATA COLLECTION PROCEDURES

1. Sample Selection:
   
a) 10% random sample of comprehensive child care licensing reviews from each licensing district, city or county.

   b) Letter with instructions is sent to district licensing/monitoring offices and/or to the child care resource and referral agencies in each child care licensing district (can be county or city-based as well).

   c) Create a cover page for each licensing office to be filled out for each licensing review, including: name of the licensing office and person collecting data; demographic information including geographic and population characteristics; and variables for separate analysis (e.g., level of staff training. See sample list of codes above).

DATA ANALYSIS

1. All analyses conducted in the pilot studies were done using the Statistical Package, SPSSPC++. State agency staff not familiar with SPSSPC++ should consult with their state information systems bureau to (a) discuss the data to be collected; (b) variables for special comparison; and (c) presentation of data.

2. If a word processing software product is used to enter the information from child care facility review reports, save the file as a DOS file. The information systems staff can enter the compliance data and design a program to accomplish the analyses you have specified.

3. It is important for state agency staff and the state information systems bureau to document the steps taken, program design, and procedures used, so that they can be easily replicated in subsequent compliance reviews.

4. If the state agency does not have the statistical support to do the data analyses and present compliance data in the formats shown on the next pages, ZERO TO THREE suggests contacting Dr. Richard Fiene, Department of Psychology, The Pennsylvania State University at Harrisburg, Middletown, Pennsylvania, 17057-4898. (see Appendix A for Dr. Fiene's other addresses and telephone numbers.) Dr. Fiene has spent much of his career developing the statistical methodologies described in each section of the manual and was the chief consultant to ZERO TO THREE's state pilot projects.
PRESENTING THE DATA

1. Table 1, *State Baseline Compliance Profile*, is the first page of a 28-page compliance profile. The table presents the cumulative percentage of non-compliance on each of that state’s 280 child care regulations. Non-compliance percentages are used to make it easier to target those regulations that should be discussed. For example, under "Policy and Procedure," state agency staff would read regulation # 006 as "55% of all child care facilities reviewed in this sample did not have a written grievance procedure for parents and staff."

2. Tables 2 and 2a, *State Compliance Profile* and *State District Compliance Report*, present a sample of data from another state and present one page of cumulative data from their comprehensive compliance review form and two pages where the data is broken down by state licensing district.

3. Both table and bar charts provide examples of how specific demographic or specific analyses can be presented. The first (Table 3) depicts the breakdown of state child care facilities by characteristic. The bar chart (Table 3a) presents the special comparison analysis relating compliance with state training requirements to total compliance — in other words, in aggregate, are child care facilities more or less likely to meet state child care regulations if staff meet state training requirements?

4. The last table, Table 4, presents a comparison of aggregate state compliance data collected at two different times, 1992 and 1993. It is clear from this table that comparing baseline to subsequent compliance studies provides state administrators data on progress in improving child care regulatory compliance as well as indicating issues needing further attention.
<table>
<thead>
<tr>
<th>Item No.</th>
<th>Compliance Requirement</th>
<th>% Out of Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Policy and Procedure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>001</td>
<td>There is a written statement of policy and procedure. The statement includes, but is not limited to the following:</td>
<td>11%</td>
</tr>
<tr>
<td>002</td>
<td>Specific philosophy and goals.</td>
<td>13%</td>
</tr>
<tr>
<td>003</td>
<td>Description of services including hours and days of operation, observed holidays.</td>
<td>13%</td>
</tr>
<tr>
<td>004</td>
<td>Population to be served.</td>
<td>15%</td>
</tr>
<tr>
<td>005</td>
<td>Fee schedules including late fees, vacation, sick day policies if applicable.</td>
<td>13%</td>
</tr>
<tr>
<td>006</td>
<td>Grievance procedures between parent(s) and director.</td>
<td>55%</td>
</tr>
<tr>
<td>007</td>
<td>Transportation policy.</td>
<td>53%</td>
</tr>
<tr>
<td><strong>Management</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>The director or qualified designee is present at the center at all times during operating hours.</td>
<td>2%</td>
</tr>
<tr>
<td>009</td>
<td>Parents are informed of the Office of Licensing. A copy of current Child Care Licensing Standards are available at the center for immediate reference by parents and staff.</td>
<td>7%</td>
</tr>
<tr>
<td>010</td>
<td>The &quot;Parent's Guide to Licensed Center Child Care&quot; is available to all parents. Appendix A</td>
<td>26%</td>
</tr>
<tr>
<td>011</td>
<td>The center is open to parents of enrolled children at all times.</td>
<td>0%</td>
</tr>
<tr>
<td>012</td>
<td>Only parents or persons authorized by parents are allowed to take any child from the center. Parents are required to authorize, in writing, persons who are allowed to take their child from the center.</td>
<td>0%</td>
</tr>
</tbody>
</table>
### TABLE 2
SAMPLE STATE COMPLIANCE PROFILE

<table>
<thead>
<tr>
<th>REGULATORY ITEMS</th>
<th>% NON-COMPLIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child Day Care Facility Inspection Checklist</td>
<td></td>
</tr>
<tr>
<td><strong>General Requirements</strong></td>
<td></td>
</tr>
<tr>
<td>1 Operator notification</td>
<td>3%</td>
</tr>
<tr>
<td>2 Change of ownership notification</td>
<td>0%</td>
</tr>
<tr>
<td>3 License displayed</td>
<td>0%</td>
</tr>
<tr>
<td>4 Licensed capacity</td>
<td>0%</td>
</tr>
<tr>
<td>5 Advertisement</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Personnel/Background Screening</strong></td>
<td></td>
</tr>
<tr>
<td>6 Household members/volunteers</td>
<td>0%</td>
</tr>
<tr>
<td>7 Abuse registry check submitted</td>
<td>11%</td>
</tr>
<tr>
<td>8 Local criminal records submission</td>
<td>11%</td>
</tr>
<tr>
<td>9 State criminal records submission</td>
<td>10%</td>
</tr>
<tr>
<td>10 Federal criminal records submission</td>
<td>10%</td>
</tr>
<tr>
<td>11 Affidavit of good moral character</td>
<td>12%</td>
</tr>
<tr>
<td><strong>Minimum Age Requirements</strong></td>
<td></td>
</tr>
<tr>
<td>12 Operator at least 21 years old</td>
<td>0%</td>
</tr>
<tr>
<td>13 Person in charge of facility</td>
<td>0%</td>
</tr>
<tr>
<td>14 16 years and older/child staff ratio</td>
<td>0%</td>
</tr>
<tr>
<td>15 Person in charge of class or group</td>
<td>0%</td>
</tr>
<tr>
<td>16 Volunteers</td>
<td>0%</td>
</tr>
<tr>
<td>17 Substitutes</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
</tr>
<tr>
<td>18 Child abuse staff/volunteer statement</td>
<td>10%</td>
</tr>
<tr>
<td>19 20 hours training</td>
<td>8%</td>
</tr>
<tr>
<td>20 8 hour annual in-service training</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Ratios</strong></td>
<td></td>
</tr>
<tr>
<td>21 Staff ratios observed</td>
<td>2%</td>
</tr>
<tr>
<td>22 Sufficient staff ratio</td>
<td>3%</td>
</tr>
<tr>
<td>23 Swimming ratio</td>
<td>1%</td>
</tr>
<tr>
<td>24 Direct supervision</td>
<td>2%</td>
</tr>
<tr>
<td>25 Substitutes available</td>
<td>1%</td>
</tr>
<tr>
<td>26 Children released to authorized individuals</td>
<td>0%</td>
</tr>
<tr>
<td>ITEMS</td>
<td>DISTRICTS (% NON-COMPLIANCE)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>General Req</td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>5</td>
</tr>
<tr>
<td>02</td>
<td>0</td>
</tr>
<tr>
<td>03</td>
<td>0</td>
</tr>
<tr>
<td>04</td>
<td>0</td>
</tr>
<tr>
<td>05</td>
<td>0</td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>0</td>
</tr>
<tr>
<td>07</td>
<td>10</td>
</tr>
<tr>
<td>08</td>
<td>10</td>
</tr>
<tr>
<td>09</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Minimum Age</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Ratios</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Physical Fac</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>10</td>
</tr>
<tr>
<td>29</td>
<td>10</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Percentage</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>21% vouchers</td>
<td></td>
</tr>
<tr>
<td>28% use subsidy</td>
<td></td>
</tr>
<tr>
<td>21% are non-profit</td>
<td></td>
</tr>
<tr>
<td>45% are profit</td>
<td></td>
</tr>
<tr>
<td>14% are religious based</td>
<td></td>
</tr>
<tr>
<td>5% are chain/franchise related</td>
<td></td>
</tr>
<tr>
<td>Average licensed capacity = 98 children</td>
<td></td>
</tr>
<tr>
<td>3% are NAEYC/NECPA accredited</td>
<td></td>
</tr>
<tr>
<td>4% are accredited by other agencies</td>
<td></td>
</tr>
<tr>
<td>59% are open 11-12 hours</td>
<td></td>
</tr>
<tr>
<td>41% are open 3-10 hours or 13+ hours</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3a
COMPARISON RELATING COMPLIANCE WITH STATE TRAINING
REQUIREMENTS TO TOTAL COMPLIANCE

SAMPLE STATE COMPLIANCE SCORES
AND STAFF HAVING A CDA

Compliance with State Child Care Regs

Staff with CDA

Staff without CDA

Compliance Scores
<table>
<thead>
<tr>
<th>Item No.</th>
<th>Compliance Requirement</th>
<th>% Out of Compliance 92</th>
<th>% Out of Compliance 93</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>There is a written statement of policy and procedure. The statement includes, but is not limited to the following:</td>
<td>11%</td>
<td>5%</td>
</tr>
<tr>
<td>002</td>
<td>Specific philosophy and goals.</td>
<td>13%</td>
<td>5%</td>
</tr>
<tr>
<td>003</td>
<td>Description of services including hours and days of operation, observed holidays.</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>004</td>
<td>Population to be served.</td>
<td>15%</td>
<td>0%</td>
</tr>
<tr>
<td>005</td>
<td>Fee schedules including late fees, vacation, sick day policies if applicable.</td>
<td>13%</td>
<td>0%</td>
</tr>
<tr>
<td>006</td>
<td>Grievance procedures between parent(s) and director.</td>
<td>55%</td>
<td>7%</td>
</tr>
<tr>
<td>007</td>
<td>Transportation policy.</td>
<td>53%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td><strong>Management</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>The director or qualified designee is present at the center at all times during operating hours.</td>
<td>2%</td>
<td>---</td>
</tr>
<tr>
<td>009</td>
<td>Parents are informed of the Office of Licensing. A copy of current Child Care Licensing Standards are available at the center for immediate reference by parents and staff.</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td>010</td>
<td>The &quot;Parent’s Guide to Licensed Center Child Care&quot; is available to all parents.</td>
<td>26%</td>
<td>3%</td>
</tr>
<tr>
<td>011</td>
<td>The center is open to parents of enrolled children at all times.</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>012</td>
<td>Only parents or persons authorized by parents are allowed to take any child from the center. Parents are required to authorize, in writing, persons who are allowed to take their child from the center.</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>
SECTION IV

Field-Tested Weighted Indicator Monitoring Tool
SECTION IV

WEIGHTED INDICATOR CHECKLIST MONITORING INSTRUMENT

PART 1: HOW EFFICIENT IS YOUR STATE’S CHILD CARE MONITORING SYSTEM?

A 1990 National Governor’s Association survey of state regulatory agencies found that only 1 out of 50 states was able to report the total number of children by age group in its regulated child care system (i.e., those programs that are licensed or registered). While many state policymakers and professionals in the field of early care and education recognize the importance of monitoring to assure compliance, local or regional monitoring agencies are chronically under-funded and short-staffed. Lack of adequate monitoring puts children at risk.

Monitoring instruments are often long and written in regulatory language, making them time-consuming to complete and difficult to understand. Monitors have large case loads, and a lengthy monitoring process leaves them little time to translate the rules or give technical assistance to programs that are out-of-compliance. Long forms and anecdotal records require time to enter into a state data base -- and often are not. Some states use short-form checklists made up by the state or local licensing office based on what they determine are the most important regulations. These may be more time-efficient and help monitors meet their caseload requirements. However, they do not predict compliance nor do they offer state and local administrators a reliable profile of child care program compliance.

The ability to collect reliable, objective data for all regulated child care programs in the state is one reason states should invest in the development of a weighted indicator checklist (WICL).
PART 2: DEVELOPING A FIELD-TESTED WEIGHTED INDICATOR CHECKLIST MONITORING TOOL

GOAL 1: TO INCREASE THE EFFICIENCY AND EFFECTIVENESS OF THE CHILD CARE MONITORING SYSTEM, GIVING MONITORS MORE TIME TO PROVIDE TECHNICAL ASSISTANCE TO HELP CHILD CARE PROGRAMS THAT DO NOT MEET CHILD CARE REGULATIONS

GOAL 2: TO INCREASE THE RELIABILITY OF THE WEIGHTED INDICATOR CHECKLIST BY SAMPLING INDIVIDUALS REPRESENTING ALL ASPECTS OF THE STATE EARLY CARE AND EDUCATION SYSTEM

A weighted indicator checklist (WICL) is a statistically reliable, field-tested monitoring instrument that can be used by child care monitors to predict compliance with the full set of child care regulations. Thus, if any item on the WICL is marked not-in-compliance, the monitor would conduct a comprehensive review of the child care facility.

A weighted indicator checklist has the following characteristics:

1) each item on the indicator checklist has been statistically tested for validity and reliability in predicting compliance;

2) field survey data and the compliance study data are incorporated into the statistical analysis;

3) each item relates to clearly observable rules thus limiting the time a monitor must spend in each child care facility in order to objectively assess compliance with child care regulations; and,

4) the weighted indicator checklist must be easily understood by both the monitors and child care providers, allowing the monitor to provide immediate feedback to child care providers on areas of non-compliance.

METHODOLOGY FOR CREATING A WEIGHTED INDICATOR CHECKLIST (WICL)

Statistical procedures, discussed in the Technical Guide (Part 3) that follows, combine data from the compliance study, discussed in Section III, and a field survey. The field survey uses a stratified sample of licensing/monitoring staff, child care professionals, parents, early care and education educators, representatives of professional organizations and child care resource and referral agencies.
The field survey asks a representative sample of 100-300 individuals, depending on the size of the state, to rank order all existing child care regulations in order of their importance for the protection of the child from risk. A Likert-like scale interval scale (with 1 = no risk and 8 = high physical or psychological risk to the child) is generally used.

Achieving a 60 to 70% response rate to the survey is important. Allow adequate time and resources for systematic follow-up.

A secondary benefit of the field survey is that it helps to build consensus in the state around the content and importance of regulation. If state policymakers want more citizen understanding and input, they can add a question that asks respondents to suggest regulations they feel are necessary but currently absent, as well as a question regarding rules that are difficult to understand or implement. That can be important information to feed into the regulatory review process.
PART 3: TECHNICAL GUIDE FOR DEVELOPING THE FIELD-TESTED WEIGHTED INDICATOR CHECKLIST

DATA ANALYSIS

1. The first step is to generate a preliminary set of weighted indicators from the data already collected by the state agency in the compliance review study described in Section III. Aggregated state compliance data are used to generate predictor items from the state’s full set of child care regulations. The SPSSPC+ program that generated the frequencies and percentages used in the compliance study must be rewritten to derive the phi coefficients.

2. The SPSSPC+ program requires recode statements in which the total compliance scores are grouped into a high-compliance group and a low-compliance group. Using the compliance review data base, which contains the compliance score for each licensed program in the statewide sample, sort the total compliance scores into a high group (top 25% in compliance) and a low group (bottom 25% in compliance). To sort the data set, use the SPSSPC+ frequency commands, and, once accomplished, run the phi coefficients through the recode command on the total compliance scores crossing this score with the total score on each regulation. SPSSPC+ generates all the phi coefficients and the respective tests of significance to determine which subset of child care rules predicts overall compliance with all state child care regulations.

3. The third data set needed to complete the state’s weighted indicator checklist is the rank ordering of child care regulations completed through the weighted field survey. Once the means are calculated, the total set of regulations can be rank ordered from highest risk value to lowest risk value. SPSSPC+ is used to combine the regulations in the high and low risk subsets with those obtained in determining the first set of predictor items. Together, they represent both predictor items from actual comprehensive child care facility records and the opinions of those from the child care field selected for the field survey.

SAMPLE SURVEY INSTRUMENTS, PRESENTATION OF FIELD SURVEY RESULTS AND A STATE WEIGHTED INDICATOR CHECKLIST

1. Child Care Weighted Field Survey: offers one format ZERO TO THREE used to present the results of the weighted field survey.

2. Sample State Child Care Monitors’ Field Survey Response: presents the results of a targeted field survey of child care monitors from two agencies (child care licensing and public health). This data was combined, aggregated and rank
ordered with the survey data from others in the child care field (child care providers, parents, professional organizations, educators). Doing this comparison allowed the state review team to examine whether there were differences in the attitudes of the two monitoring agencies.

3. **Weighted Indicator Checklist**: presents one page of items that reached significance and would appear on the state’s weighted indicator checklist. A weighted indicator checklist is usually 50 items or individual rules that if met by the child care facility predict compliance with the full set of state regulations. If any item on the weighted indicator checklist is not met by the child care facility, the monitor would go to a full compliance review.
Instructions to the reviewer: Please rate (1-8 scale) the following regulations on the basis of the risk to a child both physically and/or psychologically by being out of compliance with the specific regulation. If you have no basis for judgment of a particular item please indicate by circling 9 = "No basis for Judgment" option.

GENERAL REQUIREMENTS

1. Whenever the operator of a child care facility changes, the department or the local licensing agency must be notified in writing prior to or at the time of the change.

Please circle your response --> Low Risk <------------------> High Risk  9 = No basis for Judgment
                      1  2  3  4  5  6  7  8

2. Prior to the time a new owner assumes responsibility for a child care facility, the current owner or operator must notify the parents of the change of ownership.

Please circle your response --> Low Risk <------------------> High Risk  9 = No basis for Judgment
                      1  2  3  4  5  6  7  8

3. Upon issuance of the license it shall be displayed in a conspicuous place inside the child care facility.

Please circle your response --> Low Risk <------------------> High Risk  9 = No basis for Judgment
                      1  2  3  4  5  6  7  8

4. The child care facility must not exceed the licensed capacity designated on the license at any given time.

Please circle your response --> Low Risk <------------------> High Risk  9 = No basis for Judgment
                      1  2  3  4  5  6  7  8
<table>
<thead>
<tr>
<th>Regulation</th>
<th>Public Health</th>
<th>Licensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background screening requirements for owners, operators, employers,</td>
<td>7.00</td>
<td>7.41</td>
</tr>
<tr>
<td>volunteers (&gt;40 hrs. per mth.), family or others in contact with children--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>doesn't apply to students or persons working after hrs.; Household</td>
<td></td>
<td></td>
</tr>
<tr>
<td>members/volunteers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abuse Registry check submitted (@ yr.)</td>
<td>7.42</td>
<td>7.29</td>
</tr>
<tr>
<td>Criminal records check submission (@ 5 yrs.)</td>
<td>7.42</td>
<td>7.20</td>
</tr>
<tr>
<td>State criminal records submission (@ 5 yrs.)</td>
<td>7.25</td>
<td>7.02</td>
</tr>
<tr>
<td>Federal criminal records submission (1 time)</td>
<td>7.33</td>
<td>6.93</td>
</tr>
<tr>
<td>Affidavit of good moral character (1 time)</td>
<td>5.91</td>
<td>6.32</td>
</tr>
<tr>
<td>Minimum age requirements: Operator 21 or older</td>
<td>6.08</td>
<td>6.73</td>
</tr>
<tr>
<td>Person in charge when owner/operator absent 21 or older</td>
<td>6.27</td>
<td>6.75</td>
</tr>
<tr>
<td>Must be 16 or older to be included in ratio, must be 18</td>
<td>6.64</td>
<td>6.52</td>
</tr>
<tr>
<td>to be in charge of a group, if under 16 must be supervised, if &lt;16 cannot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>be counted in ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person in charge of a group of children 21 or older</td>
<td>6.92</td>
<td>6.95</td>
</tr>
<tr>
<td>Volunteers working &gt;40 hrs. per mth. must be 16-can be counted in ratio</td>
<td>6.73</td>
<td>6.50</td>
</tr>
<tr>
<td>Substitute considered employee-must meet age requirements above as</td>
<td>6.92</td>
<td>6.51</td>
</tr>
<tr>
<td>appropriate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training (all child care personnel except volunteers &amp; substitutes</td>
<td>6.58</td>
<td>6.35</td>
</tr>
<tr>
<td>working &lt;40 hrs. a month): Read &quot;Child Abuse &amp; Neglect in Florida, A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guide to Professionals&quot;, including volunteers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Dept. approved 20 Clock Hour Intro. Child Care Course &amp; a 10 hour</td>
<td>6.33</td>
<td>6.20</td>
</tr>
<tr>
<td>specialized training module-all</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory Items (10M-12.001-12.013)</td>
<td>PHI</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Child Day Care Facility Inspection Checklist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Abuse registry check submitted (d)(2)</td>
<td>.20133*</td>
<td></td>
</tr>
<tr>
<td>8) Local criminal records submission (d)(2)</td>
<td>.18395*</td>
<td></td>
</tr>
<tr>
<td>9) State criminal records submission (d)(2)</td>
<td>.23702*</td>
<td></td>
</tr>
<tr>
<td>10) Federal criminal records submission (d)(2)</td>
<td>.22286*</td>
<td></td>
</tr>
<tr>
<td>11) Affidavit of good moral character (e)</td>
<td>.17918*</td>
<td></td>
</tr>
<tr>
<td>18) Child abuse staff/volunteer statement (a)</td>
<td>.20899*</td>
<td></td>
</tr>
<tr>
<td>19) 20 hours training (b)(2 &amp; 3)</td>
<td>.13632**</td>
<td></td>
</tr>
<tr>
<td>22) Staff child ratios (a)(1)</td>
<td>.17020*</td>
<td></td>
</tr>
<tr>
<td>28) Dangerous supplies stored (1)(d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42) Emergency phone numbers (c)</td>
<td>.17548*</td>
<td></td>
</tr>
<tr>
<td>44) Sanitary diaper changing area (b)(2)</td>
<td>.15321**</td>
<td></td>
</tr>
<tr>
<td>45) Indoor toys/equipment safe and sanitary (a)</td>
<td>.25724*</td>
<td></td>
</tr>
<tr>
<td>46) Outdoor equipment free of hazards (b)</td>
<td>.13055**</td>
<td></td>
</tr>
<tr>
<td>56) Signed statement for discipline proc (2)(f)(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57) Personnel records maintained (3)</td>
<td>.20189*</td>
<td></td>
</tr>
<tr>
<td>59) Monthly fire drill record completed (5)(d)</td>
<td>.14520**</td>
<td></td>
</tr>
</tbody>
</table>

**Medical Inspection Checklist**

| 6) Adult on premises with current first aid (1)                                                   | .24903*   |
| 7) First aid supplies complete (2)                                                                | .19338*   |
| 12) Medical examination certificate (1)(a)                                                         | .18987**  |
| 13) Immunization requirements current (1)(b)                                                       | .16163**  |

* p < .01
** p < .05
SECTION V

The Data Display Chart and
The Recommended Action
Planning Chart
<table>
<thead>
<tr>
<th>Issue</th>
<th>Recommendations: Language adapted from APHA/AAP or panel member recommendations</th>
<th>APHA/AAP</th>
<th>0-3/NCCIP Import</th>
<th>Decisions/Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Supervision</td>
<td>Compliance reviews 1 &amp; 2 showed high compliance. Existing language is clear and observable. Might use Florida's (1) observed ratios, (2) observed adequacy of ratios, (3) adequate staffing.</td>
<td>AD9</td>
<td>*</td>
<td>** Add Rule</td>
</tr>
<tr>
<td>Ratios</td>
<td>General recommendations for best practice - scheduling staff, staff-child ratios for children with special needs, substitutes to maintain adequate staff-child ratios.</td>
<td>ST1, 2, 3 AD9 CSN37 ST60</td>
<td>*</td>
<td>** Add Rule</td>
</tr>
<tr>
<td>Group Size</td>
<td>Recommend using the National Academy of Early Childhood Programs' because of guidance on group size for mixed age groups. (ST4 specific to swimming).</td>
<td>ST1, 2, 3 AD9</td>
<td>*</td>
<td>** Add Rule</td>
</tr>
<tr>
<td>Family Child Care Director, Management Training</td>
<td>Explore management training jointly funded by CCDBG and the small Business Administration and the Department of Agriculture Extension Services. Windflower, in Colorado has Family Child Care business training. 1st create supply of appropriate training then put in role.</td>
<td>Rec.65 APPA</td>
<td>*</td>
<td>** Add Rule</td>
</tr>
<tr>
<td>Outdoor Play Area Safety (continued on next page)(see attached)</td>
<td>FA220. The following exceptions to the space requirements shall apply: (a) A minimum of 33 square feet of accessible outdoor play space is required for each infant. (b) A minimum of 50 square feet of accessible outdoor play space is required for each child from age 18 to 24 months.</td>
<td>FA220</td>
<td>*</td>
<td>** Add Rule</td>
</tr>
<tr>
<td>Regulation</td>
<td>APHA/AAP</td>
<td>Non-Compliance Rate</td>
<td>Field Rank</td>
<td>Weight Ind</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Only parents or persons authorized by parents are allowed to take any child from the center. Parents are required to authorize, in writing, persons who are allowed to take their child from the center.</td>
<td>Chapter 9 Rec 38</td>
<td>✓ ✓</td>
<td>18th</td>
<td></td>
</tr>
<tr>
<td>Persons bringing and/or picking up a child at the center sign-in and sign-out, including the time of day using a signature.</td>
<td>Chapter 9 Rec 38</td>
<td>7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caregivers sign-in, including the time of day, children who come to the center from school.</td>
<td>Chapter 9 Rec 38</td>
<td>32% 8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incidents such as injuries and medical conditions that occur while a child is in care, are reported to the parents on the day they occur.</td>
<td>HIP 88</td>
<td>✓ ✓</td>
<td>466h</td>
<td></td>
</tr>
<tr>
<td>Legal Action against a center which affects children, personnel, or operation of the center, are reported to the Department of Human Services, Office of Licensing, at the time of initiation of such legal action.</td>
<td>Chapter 9 Rec 23-25</td>
<td>✓ 2%</td>
<td>3.653</td>
<td></td>
</tr>
<tr>
<td>Center has an operable telephone. Emergency telephone numbers (fire, police, poison control, rescue unit) are posted near the telephone. Parent phone numbers are immediately available.</td>
<td>FA 1 AD 37</td>
<td>13% 7%</td>
<td>15th</td>
<td></td>
</tr>
<tr>
<td>No answering machine is used unless two or more dedicated center phone lines are available during hours of operation.</td>
<td>FA 1, AD 37 HP 105</td>
<td>27%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any person who has reason to suspect abuse or neglect reports such to an office of the Division of Family Services peace officer or law enforcement agency or regional Human Services Office.</td>
<td>HP 98</td>
<td>✓ ✓</td>
<td>40th</td>
<td></td>
</tr>
<tr>
<td>Any person whose behavior jeopardizes the health, safety and/or welfare of children or staff is not allowed to remain on the premises.</td>
<td>PR 35 (corporate, law enforcement) AP 25-35</td>
<td>✓ ✓</td>
<td>5th</td>
<td></td>
</tr>
<tr>
<td>Emergency Plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center maintains standard first aid equipment as described and recommended by the American Red Cross First Aid Handbook, current edition, or a comparable kit. Kits are stored inaccessible to children.</td>
<td>FA 116</td>
<td>24% 3%</td>
<td>33rd</td>
<td></td>
</tr>
<tr>
<td>Fire drills are conducted and documented at least monthly.</td>
<td>AD 31, AD 33, 35</td>
<td>31% 9%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ENDNOTES


REFERENCES


APPENDIX A

Resource List
APPENDIX A
RESOURCE LIST

Richard Fiene, Ph.D.
Psychologist
Bureau of Child Care
41 Grandview Avenue
Middletown, PA 17057
(717) 944-4154
(717) 944-4158 fax

Sue Aronson, M.D., FAAP, Director
Early Childhood Education Linkage System
PA Chapter of American Academy of Pediatrics
Suite 220, The Dayton Bldg.
610 Old Lancaster Road
Bryn Mawr, PA 19010-3809
(800) 24-ECELS
(610) 520-9125
(610) 520-9177 fax

Gwen Morgan
Wheelock College
The Center for Career Development in Early Care and Education
200 The Riverway
Boston, MA 02215-4176
(617) 734-5200 x210
(617) 738-0643 fax

Work/Family Directions
930 Commonwealth Avenue, West
Boston, MA 02215-1212
(617) 278-4097
(617) 566-2806 fax

Pauline Koch
(Former Chair of the National Association of Regulatory Agencies)
Division of Program Support Licensing Services
Delaware Youth & Family Center
1825 Faulkland Road
Wilmington, DE 19805-1195
(302) 633-2700
(302) 633-2652 fax

Helen Keith
1401 N. Taft Street
Suite #115
Arlington, VA 22201
(703) 522-1902

Abbey Griffin, Ph.D.
Director, Better Care for The Babies Project ZERO TO THREE/NCCIP
2000 N. 14th Street
Suite #380
Arlington, VA 22201-2500
(703) 528-4300
(703) 528-6848 fax

PENNSTATE

RICHARD FIENE
Professor in Charge
Penn State Harrisburg
Early Childhood Education
W-157 Olmsted Building
777 West Harrisburg Pike
Middletown, PA 17057-4898
Office: (717) 948-6433
Measurement Tools and Systems

Chapter 11

Richard Fiene PhD
Karen E Kroh

nara
Consumer Protection Through Prevention
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction .................................................................</td>
</tr>
<tr>
<td>Definitions ...........................................................................</td>
</tr>
<tr>
<td>Instrument Based Program Monitoring .................................</td>
</tr>
<tr>
<td>Checklists ...........................................................................</td>
</tr>
<tr>
<td>Rating Scales .......................................................................</td>
</tr>
<tr>
<td>Weighting Systems .............................................................</td>
</tr>
<tr>
<td>Licensing Indicator Systems ................................................</td>
</tr>
<tr>
<td>Outcome Based Systems ......................................................</td>
</tr>
<tr>
<td>Relationship Between Rules and Instruments ......................</td>
</tr>
<tr>
<td>Reliability and Validity ....................................................</td>
</tr>
<tr>
<td>Balance Between Compliance and Program Quality ................</td>
</tr>
<tr>
<td>Conclusion ...........................................................................</td>
</tr>
<tr>
<td>References ...........................................................................</td>
</tr>
</tbody>
</table>
PREFACE

The purpose of this chapter is to acquaint the licensing administrator with the science and art of measurement as it relates to regulatory administration. It is becoming more and more critical that licensing administrators have at least a rudimentary knowledge of measurement methods. Measurement is a key element of the new information age. It is the basis for the design and implementation of information systems, either manual or automated, conducting on-site inspections, making observations, interviewing and completing complaint investigations.

This chapter provides an overview to the major types of measurement tools used within the regulatory administration field related to assessing compliance with human care licensing rules. A historical perspective will be provided followed by outlines of key definitions. The types of measurement tools and systems will be reviewed. The final section of this chapter will address the relationship between measurement and rule formulation.

The sections titled Weighting Systems and Licensing Indicator Systems are heavily influenced by the two papers written by NARA Immediate Past President Karen E Kroh, Pennsylvania, in the late 1980s on these two topics.

Past NARA Secretary and Vice President Carolynne H Stevens, Virginia, and NARA Executive Director and Past President Pauline D Koch, Delaware, served as reviewers for this chapter.
INTRODUCTION

Measurement within regulatory administration has changed substantially from the 1970s through the 1990s. It has moved from being very qualitative to being more quantitative in nature. The qualitative nature was depicted with long narratives obtained from in-depth observations and interviews that described a facility in detail with a listing of violations with specific rules. The observations used a running record format in which a detailed accounting of the facility was obtained. This is in contrast to an anecdotal type of record that is used a great deal in the measurement literature related to observing behaviors. This qualitative system worked well when there were few facilities to be assessed. However, as the number of human care facilities increased and licensing agency administrators felt a greater need to understand compliance trends, movement to a more quantitative measurement system has evolved.

This move to quantification of measurement began in earnest in the 1970s, in particular, with the revision of the Federal Interagency Day Care Regulations (FIDCR). The notion of an instrument based program monitoring or licensing system began to be examined by licensing agencies. Checklists and rating scales were employed, with checklists being used predominantly because of the nature of regulatory compliance. However, a few states, provinces and cities utilized rating scales to measure compliance with rules. More will be said about the differences between checklists and rating scales.

By the early 1980s with severe federal cutbacks in funding, licensing administrators found themselves with an increasing number of facilities to license but fewer funds to perform the investigative function. In response to this concern, the indicator checklist methodology was created which utilized a shortened version of the comprehensive checklist approach used by many states. Indicator systems have been developing over the past two decades and in many states are key components of their monitoring and licensing functions. The indicator system is only one form of what is known in the licensing literature as inferential inspections. However, only the indicator system will be addressed in this chapter because the other types of inferential inspections are not valid and reliable enough to meet the criteria for scientifically based measurement tools.
INTRODUCTION

A related but very different technique that complements indicator systems is the use of weighting systems to determine the relative risk of specific rules related to non-compliance. The reason for the development of weighting systems is the nature of regulatory compliance data. Because compliance data measure minimum health, safety and well-being rules, the data are highly skewed with very little variance. The use of weighting systems helps to increase the amount of variance in the regulatory data sets.

The indicator and weighting systems have not been limited to licensing systems but have also been developed for other program quality endeavors such as accreditation and national standards setting.

A very recent development, in the 1990s, is the development and use of outcome based systems for licensing. This is where a licensing agency places more emphasis on outcomes rather than processes. This is a very experimental and controversial development, particularly for the field of human care licensing.
DEFINITIONS

Instrument Based Program Monitoring

A movement within licensing and regulatory administration from qualitative measurement to a very quantitative form of measurement that includes the use of checklists.

Indicator System

A licensing measurement system utilizing a shortened version of a comprehensive checklist measuring compliance with rules through a statistical methodology. Only key predictor rules are included on an indicator checklist. It is a form of inferential inspections where only a portion of the full set of rules is measured.

Inferential Inspections

An abbreviated inspection utilizing a select set of rules to be reviewed. An indicator system, weighting of rules for determining a shortened inspection tool, a random selection of rules, etc. are examples of inferential inspections. The use of inferential inspections by licensing agencies was developed as a time saving technique and a technique to focus regulatory efforts on facilities that required additional inspections or technical assistance.

Checklist

A simple measurement tool that measures compliance with state rules in a yes/no format. Either the facility is in compliance with rules or not in compliance. Generally, there is no partial compliance with checklists generally.
DEFINITIONS

Rating Scale

A more complex measurement tool in which a Likert type of rating is employed—going from more to less, or high to low. A rating scale is always used in the development of weighting systems. It is not used in measuring compliance with rules. However, rating scales are used widely in other types of program quality assessment systems—accreditation and research tools.

Weighting System

A Likert type of measurement tool that utilizes a modified Delphi technique to determine the relative risk to individuals if there are violations with specific rules. Weighting systems are developed by sending a survey to a selected sample of persons in order for them to rank the relative risk of violation with specific rules.

Outcome Based Systems

A measurement system based upon outcomes, not processes. A facility would be assessed by the outcomes it produced with individuals. For example, the number of consumers (children or adults) developing normally, free from abuse, not in placement, involved actively in the community, etc. are outcome based measures.
Instrument Based Program Monitoring (IPM) is a particular approach to measurement and assessment. It is in contrast to a more qualitative type of assessment (case study is an example of this type of assessment). IPM is very quantitative and is characterized by the use of checklists (see the next section for a discussion of checklists). The advantages of instrument based program monitoring are the following: cost savings, improved program performance, improved regulatory climate, improved information for policy and financial decisions and ability to make state/province comparisons.

IPM is a paradigm shift in conducting licensing inspections and licensing of facilities. It is an approach that lends itself to automation, it is objective and it is generally systems-oriented. The IPM approach came into its own in the 1970s and has been used predominantly since then as the primary licensing measurement tool. Some individuals have argued that the IPM approach is not as effective as the more qualitative, narrative case study approach although they can’t argue with its efficiency. A combination of IPM (quantitative approach) with a qualitative approach is probably most effective; however, this is very time consuming and a luxury that most state/province licensing agencies do not have, with more and more facilities to license and fewer and fewer staff to do the licensing.
CHECKLISTS

Checklists are the predominant means of collecting licensing data. It simplifies the process, making it very quantifiable. This is one of its strengths, but along with this simplification, a drawback is that some of the richness of the description of a particular facility is lost.

There are particular steps that need to be followed in the development of the checklist. Licensing administrators need to follow this four step process:

1) Make interpretations of the rules part of the overall manual for measurement of the comprehensive set of rules.

2) Identify the rules to be included in the checklist.

3) Consider the organization of the checklist—the flow of the investigation to the facility.

4) Decide what type of record keeping will be used—NCR paper, notebook computer in the field, etc.
RATING SCALES

Rating scales will not be discussed in detail because their applicability to licensing measurement is rather limited. Only in cases where a licensing administrator was interested in some form of partial compliance would rating scales make sense. The NAEYC (National Association for the Education of Young Children) accreditation system is one example of the use of a rating scale of full, partial or non-compliance with accreditation standards. While a partial compliance rating may be useful in accreditation standard measurement, it is generally not appropriate for use in licensing rule measurement.

Most licensing agencies do not use partial compliance, and the movement within the regulatory administration field is to consider partial compliance as being equivalent to non-compliance. Either a facility meets the rule or does not meet the rule. There is no middle ground.
WEIGHTING SYSTEMS

Weighting systems and licensing indicator systems that are described in the next section of this chapter are enhancements of the basic checklist (instrument based program monitoring) system. Weighting systems are used to increase the amount of variance in licensing compliance data. Because licensing data are nominal data (‘yes’ or ‘no’ compliance) and are generally highly in compliance, there is little variance in the data set from any particular set of rules. In order to increase the variance in data, weighting systems are used so that each rule does not have an equal weight. If you do not weight rules, by default, you have given an equal weight to each rule.

The remainder of this section describes the process for developing a licensing weighting system for use in the implementation of human care licensing rules, displays data from states that have used this approach and discusses the applicability of weighting systems for all types of human service licensing.

A licensing weighting system is a regulatory administration tool designed for use in implementing human care licensing rules. A licensing weighting system assigns a numerical score or weight to each individual licensing rule or section of a rule, based upon the relative health, safety and welfare risk to the consumers if a facility is not in compliance with the rule. The type of license issued is based on the sum of the numerical weights for each rule that is not in compliance.

The specific objectives of a licensing weighting system are:

a) To standardize decision-making about the type of license to be issued
b) To take into account the relative importance of each individual rule
c) To ensure that rules are enforced consistently
d) To improve the protection of consumers through more equitable and efficient application and enforcement of the licensing rules
A licensing weighting system can and should be developed and implemented only if:

1) Regular or full licenses are issued with less than 100% compliance with rules. If a regular license is not issued unless all violations are corrected at the time of license issuance, a weighting system is not necessary. A weighting system is useful if a facility is issued a license with outstanding violations (and a plan to correct the non-compliance areas) at the time of license issuance.

2) There is a large number of licensing rules with a variation of degrees of risk associated with various rules. If there are only a few rules with equal or similar risk associated with each rule, a weighting system is not necessary. A weighting system is useful if there are many rules with varying degrees of risk.

3) A standardized measurement system or inspection instrument is used to measure compliance with licensing rules. Before developing a weighting system, a standardized measurement instrument or tool should be developed and implemented.
WEIGHTING SYSTEMS

Development of a Weighting System

This section will provide a step-by-step process in the development of a weighting system for licensing agency use.

1) The first step in developing a licensing weighting system is the development of a survey instrument. A licensing inspection instrument or measurement tool can be adapted into a survey tool. The survey should contain each rule or section of a rule, according to how it is measured in the inspection instrument. Survey instructions should explain the purpose of the survey and instructions for completing the survey instrument. It is suggested that survey participants rate each rule section from 1-8 based on risk to the health, safety and welfare of the clients if the rule is not met (1 = least risk; 8 = most risk).

The survey participant should be instructed to circle their rating choice of 1, 2, 3, 4, 5, 6, 7 or 8. An example of a survey question is:

Interior stairways, outside steps, porches and ramps shall have well-secured handrails.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Risk</td>
<td>High Risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2) Surveys should be disseminated to at least 100 individuals. If a state has more than 3,000 licensed facilities in the type of service being surveyed, consideration for surveying more than 100 individuals should be given.

Individuals surveyed should include providers of service; provider, consumer and advocacy associations; health, sanitation, fire safety, medical, nutrition and program area professionals; licensing agency staff including policy/administrative staff and inspectors; consumers of service; and funding agency staff. In order to assure a higher survey return rate, persons selected as survey participants should be contacted prior to the survey to explain the weighting system and request their willingness to complete a survey. (See Karen Kroh’s paper for detailed graphics of Pennsylvania’s survey distribution.)
3) Survey results from each survey should be collected and entered into a computer data base spreadsheet software package. After all survey data are recorded, means or average weights for each rule or section of a rule should be calculated using SPSS—Statistical Package for the Social Sciences or SAS—Statistical Analysis System. (For detailed information on the statistical methodology employed in the development of weighting systems, see Griffin and Fiene’s *A systematic approach to child care regulatory review, policy Development of a Weighting System evaluation and planning to promote health and safety of children in child care: A manual for state and local child care and maternal and child health agency staff*.)

If there is sufficient variation in the means for each rule, the individual rule means can be rounded to the nearest whole number. Generally when comparing mean weights among the various groups surveyed there should be a similarity in rating among the groups, supporting the use of the weights as a reliable measure of risk.

4) The next step is to either (a) pilot test the weights with new licensing data for about six months or (b) apply the weights to at least 25% of historical data from the previous 12 months.

The intent of the pilot application is to collect data to use as the database for determining statistical cut-off points for the issuance of specific types of licenses or for administration of various negative sanctions.

A total weighted score for each facility based upon the combined weights of all violations should be calculated. Following is an example of how the scores should be calculated:

<table>
<thead>
<tr>
<th>RULE VIOLATIONS</th>
<th>WEIGHTS</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1</td>
<td>7</td>
</tr>
<tr>
<td># 2</td>
<td>6</td>
</tr>
<tr>
<td># 3</td>
<td>8</td>
</tr>
</tbody>
</table>

Sum of Weights = 21
WEIGHTING SYSTEMS

Under the above example a perfect compliance score with non-compliance areas would be a score of “0”. The higher the score, the lower the compliance would be. However, this is not congruent with the common usage of scores in which the higher score is associated with better compliance. In order to accommodate our familiarity with higher scores for the better facilities, the weighted score should be deducted from an arbitrary constant score of “100”. Thus a weighted non-compliance score of “20” will convert to a positive score of “80”. A facility with no violations will have a perfect score of “100”. This is more intuitive to individuals as they think about scores and measurement.

Using the previous example, the final weighted score would be computed as follows:

<table>
<thead>
<tr>
<th>RULE VIOLATIONS</th>
<th>WEIGHTS</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1</td>
<td>7</td>
</tr>
<tr>
<td># 2</td>
<td>6</td>
</tr>
<tr>
<td># 3</td>
<td>+8</td>
</tr>
</tbody>
</table>

Sum of Weights = 21

Final calculation:

\[ 100 - 21 = 79 \]
5) The fifth step in the process is to compute and apply the standard deviation or the median if the data are very skewed.

The mean and standard deviation of all final weighted scores computed in the pilot application in step #4 should then be calculated. Based upon experience with implementing licensing weighting systems, it is recommended that if a final weighted score is no more than one standard deviation below the mean, a regular license should be issued. If a score is between one standard deviation below the mean and two standard deviations below the mean, a provisional license should be issued (the length of the provisional license will vary based upon the severity of the non-compliance), or intermediate negative sanctions should be administered. If a score is less than two standard deviations below the mean, no license should be issued or a more severe negative sanction should be administered.

For example, if the standard deviation is 18 and the mean is 88, following is the distribution of the weighted scores used to determine the type of license to be issued:

- **Score of 100 — 70** = Regular license/no sanction
- **Score of 69 — 52** = Provisional license/intermediate sanction such as warnings, administrative fines or restriction on admissions
- **Score of 51 and below** = No license/severe sanction such as revocation or administrative closure

6) The final weighted scores from the pilot application should be applied to the standard deviation cut-off points to determine the type of license or negative sanction issued. These data should be studied to compare types of licenses or sanctions issued under pre-weighting vs weighting.
7) Before implementing the licensing weighting system the following additional licensing factors should be considered and incorporated as necessary into the licensing system.
   a) repeated violations from the previous licensing inspection;
   b) violation with high risk items (possibly a weight of 8.0 or above);
   c) discretion of licensing inspector to recommend variance from licensing weighting system.

8) Whenever licensing rules are amended, or at least every 5 years, the weights should be recomputed and the weighting system re-evaluated.

The licensing weighting system as described here can be used to license any type of human care facility including child care, adult care, residential care and part-day care facilities. Licensing weighting systems have been developed in Pennsylvania, Utah, Florida and Georgia.

Since the concept, development and implementation of weighting systems is relatively new to the field of licensing, the long term impact and benefits of weighting systems have not been fully realized. The potential of using weighting systems and modifications of weighting, to help standardize the implementation and enforcement of licensing rules is an exciting area of research to pursue in the field of regulatory administration.
■ LICENSING INDICATOR SYSTEMS

As mentioned in the weighting system section of this chapter, indicator checklists or licensing indicator systems are used to improve upon instrument based program monitoring (checklist) systems. The licensing indicator system is one method of assuring compliance with licensing rules in a time efficient manner. The concept has been developed and successfully implemented in several states and for different human service types. The licensing indicator system was originally developed in Pennsylvania in 1977 for use in licensing child care centers. The original intent was to develop an abbreviated licensing instrument in order to refocus licensing investigation time to assess and assist in quality enhancement activities.

From 1980-1984, the US Department of Health and Human Services funded a project to study and further develop a licensing indicator system for child day care facilities on a national level. The federally funded project, known as the Children’s Services Monitoring Transfer Consortium, organized researchers, state licensing administrators and professional staff from Pennsylvania, Michigan, West Virginia, Texas, New York City and California to review and refine the existing Pennsylvania system for possible use by other states.

The licensing indicator system is now used to assist in licensing human care facilities in Pennsylvania, West Virginia, Texas, Maryland, Utah, Florida, Delaware, Georgia, Washington, Minnesota and California.

The purpose of a licensing indicator system is to increase the efficiency and effectiveness of an existing licensing system by refocusing the emphasis of the licensing process. A licensing indicator system is intended to complement, and not replace, an existing licensing measurement system. Through use of the licensing indicator system, less time is spent conducting annual inspections of facilities with a history of high compliance with the licensing rules, and more time is spent a) providing technical assistance to help facilities comply with licensing rules and b) conducting additional inspections of facilities and agencies with low compliance with licensing rules.
The licensing indicator system is actually a shortened version of a comprehensive licensing inspection instrument. A small number of rules are selected based upon a statistical methodology designed for this specific purpose. The licensing indicator system uses a measurement tool, designed to measure compliance with a small number of rules, that predicts high compliance with all the rules. If a facility is in complete compliance with all of the rules measured in the licensing indicator system, high compliance with all the rules is statistically predicted. It is critical to understand that the rules for the licensing indicator system are selected statistically (the statistical technique is called the phi-coefficient and generally is set at a $p$ value of .01 or higher) and not based upon value judgement (arbitrary assignment, no basis from research literature), risk assessment or frequent rule violations. The rules are selected based upon an SPSSPC+ computer software package that compares violations of facilities with high compliance versus facilities with low compliance. The rules that are most often out of compliance in low compliance facilities and in compliance in high compliance facilities will be the indicator or predictor rules.

**Prerequisites for implementing a licensing indicator system**

Before developing and implementing a licensing indicator system it is important that the existing licensing system is comprehensive and well established. The following are prerequisites to implementation of an indicator system:

1) Licensing rules must be comprehensive, well written and measurable. Rules are the building blocks for any licensing system. If the rules are not well written and measurable a licensing indicator system should not be pursued. Also, if the total number of rules is small, a shortened inspection tool is not valuable.
2) There must be a measurement tool designed to standardize the application and interpretation of the rules. A licensing inspection instrument designed to assure statewide consistency in the application of the rules is essential prior to implementing a licensing indicator system.

3) There should be a licensing weighting system designed to assess the relative risk to consumers if the rule is not met. This system may be a formal weighting system or a simple classification system which categorizes rules by degree of risk. An example of a high degree of risk to consumers would be the accessibility of heat sources or toxins. Having a signature in a record is an example of a low degree of risk to consumers.

4) At least one year of data on rule violations for individual facilities. These data are needed to enter into the computer software system in order to determine the rules that are the indicators or predictors of high compliance.
How to develop a licensing indicator system

The basic steps to developing a licensing indicator system include:

1) Select facilities to be used in determining the indicators. If the total number of licensed facilities is less than 200, all 200 facilities can be used. If the total number of licensed facilities exceeds 200, sampling must be done. Generally, a sample of 100 facilities or 10% is acceptable. When selecting the sample, variables of size of facilities, geographic area, urban/rural, profit/non-profit, public/private and varied compliance levels or scores must be controlled.

2) Violation data for the sampled facilities is entered into a computer software system designed for this purpose (SPSSPC+ is recommended—consult with NARA consultant Dr. Richard Fiene for the necessary syntax and computer coding for doing the analyses).

3) A list of indicator or predictor rules, based on phi coefficients, that were the best indicators of high compliance will be calculated by the computer software system. These are the rules that are most often out of compliance in low compliance facilities and in compliance in high compliance facilities.
4) A small number of additional rules which are determined based on a licensing weighting system or relative risk are added to the statistically selected indicators. The purpose of this step is to assure face validity of the instrument. By adding a smaller number of carefully selected high-risk rules to the instrument, the licensing agency can be assured that critical rules are always measured.

5) In order to assure that full compliance with all the rules is maintained, five items selected at random should also be applied as part of the licensing indicator system. The final licensing indicator system instrument contains the indicator rules, high-risk rules and random rules. The total number of rules on an indicator checklist will vary, but will range from 20-45 items.

6) Specific criteria for use of the licensing indicator system are developed.
Criteria for use of the licensing indicator system

The development of very specific criteria for use of the licensing indicator system is perhaps the most critical step of the design process. This is the step at which the determinations are made as to when the licensing indicator system will be used. The determination of use of the system should be standardized and not based upon licensing inspector discretion.

Each licensing agency must develop its own criteria based upon its own historical licensing data and experience. Following are some criteria that may be useful:

1) The facility has had a full or regular license and no negative sanctions have been administered, within the previous two (2) years.

2) The facility has had a score or percentage of compliance above a specified threshold for the previous year.

3) All previous violations have been corrected according to the facility’s plan of correction.

4) No significant validated complaints have been found within the past year.

5) The total number of consumers served has not increased by more than a specified percentage during the past year.

6) There has not been significant staff turnover at the facility/agency within the past year. This may be targeted to certain levels of staff turnover, such as direct care staff or facility directors, depending on which staff are particularly key for program stability.

7) A full inspection using the comprehensive licensing measurement instrument must be done at least every three (3) years.
LICENSING INDICATOR SYSTEMS

Revision of the licensing indicator system

The licensing indicator system should be continually reevaluated for its effectiveness. The system should be completely revised at least every three years or upon a revision of the rules. In order to achieve the intended purpose of the licensing indicator system of refocusing the emphasis of licensing effort from facilities with high compliance to facilities with low compliance, constant review, evaluation and revision of the licensing indicator system is essential.

Other types of inferential inspection systems, of which the licensing indicator system is only one, will not be addressed in this chapter because inferential systems other than the licensing indicator system have not been determined to be statistically valid or reliable. As licensing administrators may potentially need to defend their actions in a court of law, it is essential that the methodology or technique utilized is scientifically sound. When it comes to inferential inspections only those instruments based upon an indicator or weighting methodology can stand up to this rigorous testing.
OUTCOME BASED SYSTEMS

This is a relatively new phenomenon in the licensing and regulatory administration field. The emphasis in this new approach is to examine outcomes rather than processes. What are the ultimate outcomes for individuals? Determine this and the argument goes—there is no need to measure processes directly.

Outcome measurement is appealing in many respects. It does focus on results, something the human services field was short on demonstrating in the 1990s. However, there is a fallacy in this approach. Results are the end product, but we always have a process to get to the end product.

Another issue is that the purpose of licensing is to prevent harm to consumers. A purely outcome-based system would potentially harm consumers who were in the facilities later determined to “fail” the outcome test. Moreover, there are two other problems:

1. Insufficient (political) agreement on what are acceptable outcomes.
2. Some outcomes will not manifest for years and/or are contaminated by other variables related to other influences on later behavior.

What makes more sense is to tie outcomes to specific regulatory processes that appear to be in a causal or at least a correlational relationship. If licensing agencies were able to clearly link specific results (outcomes) to specific rules (processes), there would be the empirical ability to focus only on those rules that produced positive results for consumers and families and eliminate all other unnecessary rules that do not produce positive outcomes for consumers and families. Specific studies could be conducted and in fact have already been conducted by university researchers. In child care, for example, low staff:child ratios, pre-service and in-service training of staff, highly qualified staff and small group size are all examples of regulatory variables that have been identified as surrogates to program quality that produce positive outcomes for children.

Outcome based or results-oriented systems will impact licensing, but the research literature demonstrates how licensing agencies can clearly link outcomes to regulatory processes that produce the outcomes. This becomes a powerful argument to legislators when this roadmap of process to outcome can be provided.
RELATIONSHIP BETWEEN RULES AND INSTRUMENTS

This section is included because this is one area that gets many licensing administrators into trouble. Not enough time is spent on making sure that the instruments developed are the exact reflection of the rules. This is where the interpretive rules that are part of any measurement instrument that accompanies the actual instrument should be placed. This helps to increase the reliability of the instrument and doesn’t hurt the overall validity of the tool either (more on reliability and validity in the next section). Readers should refer to Chapter 2, The Formulation of Rules, for additional information on the definition and development of interpretive and substantive rules.

When there is not a close link between instrument development and rule formulation this only leads to headaches for licensing agencies. It may take years and not be evident until you get called into a court of law to defend your licensing system but it will happen.

The analogy of playing Russian Roulette may be useful. As licensing administrators, you are never 100% certain that all your facilities are compliant with all the rules. However, there are certain management procedures and processes that you can put in place to help. A clear link between rules and measurement tools is one of them. Since you are never 100% sure of full compliance (in other words all six chambers of the revolver are not empty—if they were, you wouldn’t have Russian Roulette), you must make difficult decisions related to increasing or decreasing your chances in playing Russian Roulette. So you have the choice of having the management and procedural safeguards built in (one or two bullets in the revolver) or you don’t build in the procedural safeguards (four or five bullets in the revolver). It is obvious statistically where your chances are greater in surviving a potential mishap in a licensing system.


RELIABILITY AND VALIDITY

The two concepts of reliability and validity are so critical to measurement, but are so often overlooked in the development of licensing measurement systems. In fact, it has been estimated that as many as 30 states may be using a type of inferential inspection. But only 1/3 of these states has followed the rigorous statistical methodology as outlined in the Licensing Indicator System section.

Validity and Reliability

Very simply, validity deals with content of the particular tool or instrument—does it serve the purpose for which it is to be used? Does it measure the rules accurately? Usually the answer to this question is easier for licensing administrators to answer. Since licensing measurement tools should be directly based upon rules, as explained in the previous section, there should not be much difficulty in establishing validity. When the tools are not based on the rules, that is when validity can be and should be called into question.
**RELIABILITY AND VALIDITY**

**Reliability** deals with the administration of the tool or instrument. Does it measure the rules consistently and in an objective manner? The answer to this question is much more difficult for licensing administrators to answer affirmatively. This poses real problems if each administration of the licensing tool is not consistent and objective. Facilities will not have the rules applied in an equal and fair manner.

Reliability testing should be done methodologically and scientifically. Interrater reliability should be established for the tools/checklists that are to be used in the field by licensing field staff. This is a process that has been well documented in the psychological research. This has not been the case within licensing and regulatory administration. Generally checklists are designed quickly and are never tested for reliability. This creates a problem that many of us have heard—the rules are not applied uniformly across the state/province. The reason is that the tool that is used to measure compliance is not reliable.

In order to establish reliability, licensing inspectors need to go out to facilities in pairs assessing the same facility at the same time. They then need to compare their results. Do they agree on what is in compliance and out of compliance at the particular facility? If there is not at least 90% agreement for each rule then additional interpretation of that specific rule is needed. Establishing reliability is not overly difficult nor overly time consuming; however, it will add a bit more time before staff are really ready to begin to license facilities (90% agreement on each rule and interpretative rule).
BALANCE BETWEEN COMPLIANCE AND PROGRAM QUALITY

An interesting development in the past five years has been the emphasis on program quality as a result of pressure from consumers, families, advocates and the general public. Consumers and other interested persons are requesting licensing agencies to ensure not only the health, safety and well-being of individuals served in facilities, but also to be concerned advocates for the overall quality of services provided at these facilities.

This increased emphasis and concern for program quality is a difficult area to address for licensing agencies. The resources to complete program quality reviews and to advocate for quality within government are not commensurate with the expectations. However, there are some strategies that can be employed to assist licensing agencies. The first and foremost will be to save time on doing licensing inspections. The indicator system described in this chapter will provide such a tool for saving time. Studies conducted over the past two decades indicate that utilizing an indicator checklist approach saves up to 50% in the on-site inspection time.

The time saved in doing licensing inspections should be used to either:

   a) Conduct additional licensing inspections in new or problem facilities
   b) Provide technical assistance
   c) Complete program quality reviews

This could be done by utilizing a tool from accreditation in observing classrooms, or utilizing a program quality tool from the research literature (for example, Early Childhood Environment Rating Scale). Licensing administrators need to be certain that they have a plan to utilize this extra time or the worst fears of licensing professionals could occur. Two potential scenarios could play out. One is that the time is used to do more and more licensing inspections utilizing the indicator system on more and more facilities. The worst scenario is that staffs are cut. If a state/province can complete all its inspections in half the time, then doesn’t it follow that only half the staff is needed? With a clearly articulated plan on how the licensing and program quality reviews will produce higher quality programs should help to prevent this cost cutting approach. However, this is always a fear that licensing administrators must face.
CONCLUSION

This NARA Licensing Curriculum chapter provides a brief overview to the major issues confronting licensing administrators when they consider licensing tools and measurement systems. The emphasis upon quantitative systems was reflected in this chapter because of the need to develop cost effective and efficient licensing systems as the number of facilities continues to grow with shrinking resources. Also there is a compounding effect with higher expectations on licensing agencies to be concerned more about program quality.

The chapter showed the various types of measurement tools that apply to licensing and regulatory administration. It is clear that given the nature of licensing there are certain tools more suited than others, such as checklists versus rating scales. A very detailed description of both licensing weighting and indicator systems was provided. The reason for this emphasis is that these are two very valid and reliable tools that can be used by licensing administrators in making their agencies more effective and efficient. The licensing measurement field is changing constantly as new approaches are introduced. For example, within the program evaluation field there is a move to have a better balance between quantitative and qualitative analyses. It will not be long before this initiative has its impact on the licensing measurement field as well.
REFERENCES


Introduction

The purpose of this paper is to compare several countries (N = 20) and the United States on the Child Care Aware – formerly NACCRA (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; NACCRRRA 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:*

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

![Figure 1](image-url)
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 - NACCRRRA 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 - NACCRRRA 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 - NACCRRRA Benchmarks; 0.5 = Marginally meets the Child Care Aware - NACCRRRA Benchmarks; 1.0 = Partially meets the Child Care Aware - NACCRRRA Benchmarks; 1.5 = Substantially meets the Child Care Aware - NACCRRRA Benchmarks.
Benchmarks; 2.0 = Fully meets the Child Care Aware – NACCRAA 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 – NACCRAA 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
Table 1

*Mean Comparisons between USA and Twenty Countries on Child Care Aware – NACCRRA Benchmarks*

<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><strong>1.5625</strong></td>
<td>0.5</td>
<td>t = 7.100; p &lt; .0001</td>
</tr>
<tr>
<td>Teacher (R4)</td>
<td><strong>1.6563</strong></td>
<td>0.4038</td>
<td>t = 7.632; p &lt; .0001</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><strong>1.0481</strong></td>
<td>t = 2.534; p &lt; .02</td>
</tr>
<tr>
<td>Clearances (R7)</td>
<td>0.6094</td>
<td><strong>1.2404</strong></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><strong>1.7404</strong></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.

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.
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 NAEC Accreditation Standards met (R2)
ACR = Staff child ratios NAEC Accreditation Standards met (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
A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA Child Care 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 Podschlne, 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: NACCRRA Ranking of State Child Care Center Regulations: 2011 Update* were developed by Child Care Aware - NACCRRA 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.

<table>
<thead>
<tr>
<th>Scoring Methods for NACCRRA Ranking of State Child Care Center Regulations (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question</strong></td>
</tr>
<tr>
<td><strong>Scoring method</strong></td>
</tr>
<tr>
<td><strong>Regulation 1. Staff:child ratio requirements comply with NAEYC accreditation standards.</strong></td>
</tr>
<tr>
<td>Number of ratios in compliance with NAEYC standards</td>
</tr>
<tr>
<td>Score</td>
</tr>
<tr>
<td>7 ratios</td>
</tr>
<tr>
<td>6 ratios</td>
</tr>
<tr>
<td>5 ratios</td>
</tr>
<tr>
<td>4 ratios</td>
</tr>
<tr>
<td>3 ratios</td>
</tr>
<tr>
<td>2 ratios</td>
</tr>
<tr>
<td>1 ratios</td>
</tr>
<tr>
<td><strong>R2. Group size requirements are in compliance with NAEYC accreditation standards.</strong></td>
</tr>
<tr>
<td>Number of group sizes in compliance with NAEYC standards</td>
</tr>
<tr>
<td>Score</td>
</tr>
<tr>
<td>7 ratios</td>
</tr>
<tr>
<td>6 ratios</td>
</tr>
<tr>
<td>5 ratios</td>
</tr>
<tr>
<td>4 ratios</td>
</tr>
<tr>
<td>3 ratios</td>
</tr>
<tr>
<td>2 ratios</td>
</tr>
<tr>
<td>1 ratios</td>
</tr>
</tbody>
</table>
**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 ≥</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>
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 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

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

* Corresponding author. Tel.: +520-326-5154; fax: +520-326-5155.
E-mail addresses: mlahti@usm.maine.edu, Michel.Lahti@lecroymilligan.com (M. Lahti).
1 Professor Emeritus at Pennsylvania State University.

http://dx.doi.org/10.1016/j.ecresq.2014.04.005
0885-2006/© 2014 Elsevier Inc. All rights reserved.

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
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 (Cizek, 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 design decisions 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, ECEC 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
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 “right” ones to include by examining levels of empirical and expert support.</td>
<td><em>Do the quality components capture the key elements of quality?</em></td>
<td><em>Process subject to interpretation and to political pressure</em></td>
</tr>
<tr>
<td>2. Examine the measurement strategy and psychometric properties of measures used to assess quality.</td>
<td><em>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 psychometrically sound.</em> <em>Examine relationships among quality measures to assess whether they function as expected.</em></td>
<td><em>Is there sufficient empirical and expert support for including each standard?</em> <em>What is the reliability and accuracy of indicators collected using different methods?</em></td>
<td><em>Limited empirical evidence available; few established links to outcomes of interest.</em> <em>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.</em></td>
</tr>
<tr>
<td>3. Assess the outputs of the rating process</td>
<td><em>Examine variation and patterns of program-level ratings within and across program types, to assess if QRIS distinguishes levels of quality.</em> <em>Examine relationship of program-level ratings to other validated quality indicators to determine if ratings are assessing quality in expected ways.</em> <em>Examine alternate cut points and combining rules to determine how well the ratings distinguish different levels of quality.</em></td>
<td><em>Do rating distributions vary by program type, e.g., center-based programs vs. home-based programs?</em></td>
<td><em>Measurement error is an important issue that should be examined.</em></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><em>Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?</em> <em>Do levels cut scores and combining rules produce expected rating distributions and meaningful distinctions among programs?</em></td>
<td><em>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.</em></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.

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 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 CCDF/Voluntary for all others</td>
</tr>
<tr>
<td>QRIS structure</td>
<td>Building Block; 4 quality levels</td>
<td>Building Block; 4 quality levels</td>
</tr>
<tr>
<td>QRIS standards (examples)</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 development process</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.html">www.in.gov/fssa/carefinder/2554.html</a>).</td>
<td>Step 4 – National accreditation; written parent involvement plan; etc. (See: <a href="https://www.maine.gov/dhhs/oecs/ecs/in/qualityforms.htm">https://www.maine.gov/dhhs/oecs/ecs/in/qualityforms.htm</a>).</td>
</tr>
<tr>
<td>QRIS rating procedure</td>
<td>Independent ratings contractor, annual site visits, using Paths to QUALITY standards checklist</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>
</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, & Zaslavow, 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 Fiene (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, Hall, & 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.

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
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 4.3, 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 QRIS 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.

Examinations of 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 ministries, 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...
Table 3
Indiana QRIS: mean global quality ERS scores* 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 (.84)</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>

* Possible range = 1–7.

Table 4
Maine QRIS: mean global quality ERS scores* 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>

* Possible range = 1–7.

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 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 family 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 "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 highest 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 .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
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 (Zaslow, Martinez-Beck, Tout, & Halle, 2011).

Maine: assessing the input 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 home scores 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, and 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.

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
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 ($r = .17^{**}$) and 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 correlates 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 or Indiana and Maine). Use of these approaches enables cross system comparisons.

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
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 appear 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


INTRODUCTION

The purpose of this report is to provide the Kansas Child Care Office with basic analyses for the development of their key indicator system for both centers and homes. Licensing data from 2012 taken from both centers (CCC) (n = 482) and homes (FCC) (n = 500) were used in this Licensing Key Indicator study. The centers were further broken down into 52 (11%) Head Start programs and 430 (89%) child care centers. The homes were further broken down into 115 (23%) group homes and 385 (77%) family homes.

Definitions:

Key Indicators (KI) = a differential monitoring approach that employs using only those rules that statistically predict overall compliance with all the rules. In other words, if a program is 100% in compliance with the Key Indicators the program will also be in substantial to full compliance with all rules. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators, the program will also have other areas of non-compliance with all the rules. In this study, 8 Key Indicator rules were identified for CCC and 6 Key Indicator rules for FCC. The Key Indicators can be found in the Findings Section of this report.

Rule Violations or Citations = this occurs when a program does not meet a specific rule and is cited as being out of compliance with that rule.

METHODOLOGY

A Differential Monitoring Logic Model and Algorithm (DMLMA©)(Fiene, 2012) was employed, in particular, the key indicator methodology to generate the Key Indicators for this project. The DMLMA© is the 4th generation of an Early Childhood Program Quality Indicator Model (ECPQIM)(Fiene & Nixon, 1985; Griffin & Fiene, 1995; Fiene & Kroh, 2000).

The DMLMA© (see Figure 1) provides the conceptual model for assessing the overall effectiveness of a differential monitoring system. The two main tools in a Differential Monitoring (DM) system are Risk Assessment (RA) and Key Indicator (KI) measurement tools. Both the Risk Assessment and Key Indicator tools are derived from a comprehensive licensing tool (CI) that measures compliance with all rules. For the purposes of this study the Licensing Data taken from Kansas Monitoring Reviews represents the comprehensive licensing tool (CI). Kansas presently does not use a Risk Assessment or a Program Quality tool (see Table 1).

Table 1

<table>
<thead>
<tr>
<th>DMLMA© Terminology</th>
<th>Kansas Examples and Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive Tool (CI)</td>
<td>Licensing Data from Kansas Monitoring Visits</td>
</tr>
<tr>
<td>Program Quality Tool (PQ)</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Risk Assessment Tool (RA)</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Key Indicators (KI)</td>
<td>Generated from this Study</td>
</tr>
<tr>
<td>Differential Monitoring (DM)</td>
<td>Not Applicable</td>
</tr>
</tbody>
</table>
FINDINGS

There are some overall demographic findings presented first that help to put the results in context. As mentioned in the introduction there were 482 centers and 500 homes that were part of these analyses. Eleven percent (11%) of the centers were 100% in compliance with all rules while 25% of the homes were 100% in compliance with all rules. These figures are fairly typical of state averages. The average number of violations for centers was 7.44 violations with all applicable rules and 3.52 violations for homes.

Location of the various facilities seemed to have an impact on average violations recorded. For example, with centers, urban facilities had a significantly higher level of violations (8.42 average violations; n = 279) than facilities located in rural communities (6.09 average violations; n = 203). This result was statistically significant (F = 14.19; p < .0001). However, the differences for homes was not statistically significant, with urban homes (n = 222) having 3.64 average violations versus 3.42 average violations for rural homes (n = 278).

There were statistically significant differences depending on the Region the facilities were located in. For centers, the highest levels of violations with child care rules were in Regions 1 (9.30 average violations; n = 109) and 2 (8.32 average violations; n = 191) while Regions 3 (5.31 average violations; n = 121) and 4 (5.57 average violations; n = 61) had lower averages (see Table 2). This result is statistically significant (F = 9.82; p < .0001).

Table 2: Violation Data in Centers and Homes by Regional Location

<table>
<thead>
<tr>
<th>Region</th>
<th>Centers</th>
<th></th>
<th></th>
<th>Homes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Violations*</td>
<td>Number</td>
<td>Violations*</td>
<td>Number</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.30</td>
<td>109</td>
<td>2.42</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8.32</td>
<td>191</td>
<td>4.63</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.31</td>
<td>121</td>
<td>3.94</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.57</td>
<td>61</td>
<td>3.02</td>
<td>125</td>
<td></td>
</tr>
</tbody>
</table>

* = Average Violations (Mean)

For homes, a slightly different distribution occurs in which Region 2 (4.63 average violations; n = 120) was significantly higher than the other three regions. This result is statistically significant (F = 7.24; p < .0001).

Also the type of licensing inspection saw some variation in the average number of violations although none of the following results were statistically significant (see Table 3).

Table 3: Violation Data in Centers and Homes by Type of Licensing Inspection

<table>
<thead>
<tr>
<th>License Type</th>
<th>Centers</th>
<th></th>
<th></th>
<th>Homes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Violations*</td>
<td>Number</td>
<td>Violations*</td>
<td>Number</td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>7.44</td>
<td>36</td>
<td>3.35</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Renewal</td>
<td>7.07</td>
<td>368</td>
<td>3.53</td>
<td>469</td>
<td></td>
</tr>
<tr>
<td>Amendment</td>
<td>9.51</td>
<td>55</td>
<td>4.00</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Correction</td>
<td>6.71</td>
<td>14</td>
<td>3.00</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Temporary</td>
<td>11.22</td>
<td>9</td>
<td>4.00</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

* = Average Violations (Mean)

The last demographic analysis was to compare the average number of violations between group homes and family homes; and between child care centers and Head Start programs. There was not a significant difference between group homes (3.75 average violations; n = 115) and family homes (3.45 average violations; n = 385); but a statistically significant difference occurred (F = 10.44; p < .001) between child care centers (7.78 average violations; n = 430) and Head Start programs (4.60 average violations; n = 52) with the Head Start programs having significantly fewer rule violations.
Key Indicator Findings

The following findings will provide the Key Indicators for centers (child care centers and Head Start) and homes (family and group homes). It will provide a listing of the rules and the respective phi coefficients. These Key Indicators were obtained from rank ordering the total compliance scores into quartiles with the 25% highest violation scores for facilities as the low group and the lowest 25% violation scores for facilities as the high group. Each rule was compared to this result by their respective compliance level, either being in or out of compliance with the rule. Once these data were prepared the formula in Table 4 was used to determine if the rule met the predictive level. Separate analyses for generating Key Indicators were not run for Head Start or Group Homes because of the insufficient number of programs in each category.

Centers (Child Care Centers and Head Start)(See Table 5 for a Summary)

All results are reported with the specific rule, p < .0001, and phi coefficient from the formula in Table 4.

K.A.R.28-4-126b1. Each person regularly caring for children shall have a health assessment conducted by a licensed physician or by a nurse trained to perform health assessments. The health assessment shall be conducted no earlier than one year before the date of employment or initial application for a license or certificate of registration, or not later than 30 days after the date of employment or initial application. (phi = .59)

K.A.R.28-4-126c1. Each person living, working or regularly volunteering in the facility shall have a record of a negative tuberculin test or x-ray obtained not more than two years before the employment or initial application, for a license or certificate of registration or not later than 30 days after the date of employment or initial application. (phi = .62)

K.A.R.28-4-423a18. The premises shall be maintained in good condition and shall be clean at all times, free from accumulated dirt and trash, and any evidence of vermin or rodent infestation. Each outdoor trash and garbage container shall be covered, and the contents shall be removed at least weekly. (phi = .59)

K.A.R.28-4-423a23. Medicines, household poisons, and other dangerous substances and instruments shall be in locked storage. (phi = .60)

K.A.R.28-4-428aa3. Each licensee shall ensure that orientation is completed by each staff member who will be counted in the staff-child ratio and by each volunteer who will be counted in the staff-child ratio. Each staff member and volunteer shall complete the orientation within seven calendar days after the date of employment or volunteering and before the staff member or volunteer is given sole responsibility for the care and supervision of children. (phi = .51)

K.A.R.28-4-428ac1. Each staff member counted in the staff-child ratio, each volunteer counted in the staff-child ratio, and each program director shall obtain certification in pediatric first aid and in pediatric CPR as specified in this subsection either before the date of employment or volunteering or not later than 30 calendar days after the date of employment or volunteering. (phi = .53)

K.A.R.28-4-430c3. Each staff member shall be trained to observe symptoms of illness, neglect, and child abuse, and shall observe each child’s physical condition daily. (phi = .54)

K.A.R.28-4-437d. The outdoor play space shall be well drained and free of hazards. (phi = .59)

Footnote:
Child Care Centers (CCC) – The correlation between the Key Indicators and all the rules was .77.
Family Child Care (FCC) – The correlation between the Key Indicators and all the rules was .80.
Both these results exceed the DMLMA© Thresholds for KI x CI (.70).
Homes (Family and Group Homes)(See Table 5 for a Summary)

All results are reported with the specific rule, p < .0001, and phi coefficient from the formula in Table 4.

K.A.R.28-4-115g1. All household cleaning supplies and all bodily care products bearing warning labels to keep out of reach of children or containing alcohol shall be in locked storage or stored out of reach of children under six years of age. Soap used for hand washing may be kept unlocked and placed on the back of the counter by a bathroom or kitchen sink.  (\(\text{phi} = .47\))

K.A.R.28-4-115aa1A. Supervision plan. Each applicant, each applicant with a temporary permit, and each licensee shall develop a supervision plan for children in care that includes all age ranges of children for whom care will be provided. A copy of the plan shall be available for review by the parents or legal guardians of children in care and by the department. The plan shall include the following: A description of the rooms, levels, or areas of the facility including indoor and outdoor areas in which the child will participate in activities, have snacks or meals, nap, or sleep.  (\(\text{phi} = .79\))

K.A.R.28-4-115aa1B. Supervision plan. Each applicant, each applicant with a temporary permit, and each licensee shall develop a supervision plan for children in care that includes all age ranges of children for whom care will be provided. A copy of the plan shall be available for review by the parents or legal guardians of children in care and by the department. The plan shall include the following: the manner in which supervision will be provided.  (\(\text{phi} = .44\))

K.A.R.28-4-117a1. A completed medical record on a form supplied by the department shall be on file for each child under 11 years of age enrolled for care and for each child under 16 living in the child care facility.  (\(\text{phi} = .44\))

K.A.R.28-4-117c. Immunizations for each child, including each child of the provider under 16 years of age shall be current as medically appropriate and shall be maintained current for protection from the diseases specified in K.A.R. 28-1-20(d). A record of each child's immunizations shall be maintained on the child's medical record.  (\(\text{phi} = .68\))

K.A.R.28-4-127b1A. Emergency medical treatment: Each facility shall have on file at the facility for each child: written permission of the parent, guardian, or legal custodian for emergency medical treatment on a form that meets the requirements of the hospital or clinic where emergency medical care will be given.  (\(\text{phi} = .53\))

References


### Table 4: Kansas Key Indicator (KSKI) Formula Matrix

<table>
<thead>
<tr>
<th></th>
<th>Providers In Compliance</th>
<th>Programs Out Of Compliance</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Group A</td>
<td></td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Low Group C</td>
<td></td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td>Column Total W</td>
<td>W</td>
<td>X</td>
<td>Grand Total</td>
</tr>
</tbody>
</table>

**Key Indicator Statistical Methodology (Calculating the Phi Coefficient):**

\[
\phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}}
\]

- \(A = \text{High Group + Programs in Compliance on Specific Compliance Measure.}\)
- \(B = \text{High Group + Programs out of Compliance on Specific Compliance Measure.}\)
- \(C = \text{Low Group + Programs in Compliance on Specific Compliance Measure.}\)
- \(D = \text{Low Group + Programs out of Compliance on Specific Compliance Measure.}\)
- \(W = \text{Total Number of Programs in Compliance on Specific Compliance Measure.}\)
- \(X = \text{Total Number of Programs out of Compliance on Specific Compliance Measure.}\)
- \(Y = \text{Total Number of Programs in High Group.}\)
- \(Z = \text{Total Number of Programs in Low Group.}\)

**Phi Coefficient Range**

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include on KSKI</td>
</tr>
<tr>
<td>(.25) – (0)</td>
<td>Too Easy</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(0) – (-.25)</td>
<td>Too Difficult</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>

*High Group = Top 25% of Programs in Compliance with all Compliance Measures.*

*Low Group = Bottom 25% of Programs in Compliance with all Compliance Measures.*
FIGURE 1- DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (Fiene, 2012)

*DMLMA*© Applied to the Kansas Child Care Licensing System

\[ CI + PQ \Rightarrow RA + KI \Rightarrow DM \]

**Kansas Examples:**

- **CI** = Licensing Reviews (All Rules)
- **PQ** = Not Applicable (NA)
- **RA** = Not Applicable (NA)
- **KI** = Key Indicators (generated from this study)
- **DM** = Not Applicable (NA)

*DMLMA*© *Thresholds:*

- **High Correlations (.70+)** = CI x KI
- **Moderate Correlations (.50+)** = CI x RA; RA x DM; RA x KI; KI x DM
- **Lower Correlations (.30+)** = PQ x CI; PQ x RA; PQ x KI
Table 5 – Rule Numbers and Phi Coefficients for Centers and Homes

<table>
<thead>
<tr>
<th>Centers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule</td>
<td>Phi</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-126b1.</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-126c1.</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-423a18.</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-423a23.</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-428aa3.</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-428ae1.</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-430c3.</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-437d.</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Homes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule</td>
<td>Phi</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-115g1.</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-115aa1A.</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-115aa1B.</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-117a1.</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-117c.</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>K.A.R.28-4-127b1A.</td>
<td>.53</td>
<td></td>
</tr>
</tbody>
</table>

For additional information regarding this report, please contact:
Richard Fiene, Ph.D., Director
Research Institute for Key Indicators
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
ResearchInstituteKeyIndicators@ymail.com
717-944-5868 Phone and Fax
http://pennstate.academia.edu/RickFiene
Georgia Child Care Licensing Study: Validating the Core Rule Differential Monitoring System

Executive Summary

Richard Fiene, Ph.D.

The purpose of this study was to validate Georgia’s process for determining if a state-regulated child care facility is compliant with basic state health and safety requirements. The process was developed by staff at Bright from the Start: Georgia Department of Early Care and Learning (DECAL). Currently Georgia utilizes a “Core Rule” risk assessment approach in which the health and safety rules deemed most crucial to ensure children’s health and safety are used to compute a program’s compliance status.

This validation study utilized a unique analytical model that compared licensing data with previous key indicator (for readers not familiar with this term, please see the definitions on page 4 of the report) research and ascertained if the Core Rules accurately indicated a program’s overall compliance with the total population of licensing rules.

Additional statistical analyses examined if the mathematical formula used to compute compliance was an appropriate configuration of the data that discerned between those programs that adequately met basic health and safety rules (compliant) and those that did not (non-compliant). Also licensing data were compared to a representative sample of quality data collected as part of a different study to examine the correlation between compliance and quality. A Differential Monitoring Logic Model/Algorithm (DMLMA©) (Fiene, 2012) and a previous validation model (Zellman & Fiene, 2012) were used in the research.

One hundred and four child care centers (104 CCC) and 147 family child care (FCC) homes were assessed. Licensing data over a four-year period (2008-2012) and matching program quality data from a two-year period (2007-2008) were used in this study.

The study focused on three research questions:

1. Do the Core Rules CCCs and FCC homes serve as overall Key Indicators of compliance?
2. Does the Annual Compliance Determination Worksheet (ACDW) appropriately designate programs as compliant or non-compliant related to health and safety?
3. Are the Core Rules related to program quality?

The analysis demonstrated that the Core Rules did serve as key indicators, and these key indicators were identified for both center based and home based child care. The second analysis concluded that the ACDW computation did distinguish between compliant and non-compliant programs. Finally, the expected correlation between compliance and quality was found but only for state-funded Pre-K classrooms, not for family child care nor for preschool classrooms that were not part of the state-funded Pre-K.
Georgia Child Care Licensing Study: Validating the Core Rule Differential Monitoring System

Richard Fiene, Ph.D.

February 1, 2014

This study was made possible by a grant from Bright from the Start: Georgia Department of Early Care and Learning. All opinions expressed in the report reflect the opinions of the author, not necessarily those of the Department of Early Care and Learning.

ABSTRACT

The purpose of this study was to validate Georgia’s process for determining if a state-regulated child care facility is compliant with basic state health and safety requirements. The process was developed by staff at Bright from the Start: Georgia Department of Early Care and Learning (DECAL). Currently Georgia utilizes a “Core Rule” risk assessment approach in which the health and safety rules deemed most crucial to ensure children’s health and safety are used to compute a program’s compliance status. This validation study utilized a unique analytical model that compared licensing data with previous key indicator (for readers not familiar with this term, please see the definitions on page 4 of the report) research and ascertained if the Core Rules accurately indicated a program’s overall compliance with the total population of licensing rules. Additional statistical analyses examined if the mathematical formula used to compute compliance was an appropriate configuration of the data that discerned between those programs that adequately met basic health and safety rules (compliant) and those that did not (non-compliant). Also licensing data were compared to a representative sample of quality data collected as part of a different study to examine the correlation between compliance and quality. A Differential Monitoring Logic Model/Algorithm (DMLMA©) (Fiene, 2012) and a previous validation model (Zellman & Fiene, 2012) were used in the research. Child care centers (CCC) and family child care (FCC) homes were assessed. The analysis demonstrated that the Core Rules did serve as key indicators, though this list should be reexamined. The second analysis concluded that the computation could be simplified. Finally, the expected correlation between compliance and quality was found but only in state-funded Pre-K classrooms; it was not found in preschool classrooms and could not be validated. Family child care could not be validated either. As a result of the study, recommendations were made to strengthen Georgia’s system.

Acknowledgements:

Special thanks are extended to DECAL staff who had the vision to conduct this validation study: Bobby Cagle, Commissioner; Kay Hellwig, Assistant Commissioner for Child Care Services; Kristie Lewis, Director of Child Care Services; and Dr. Bentley Ponder, Director of Research & Evaluation. Also, researchers at the University of North Carolina, Chapel Hill, Frank Porter Graham Child Development Institute, Dr. Donna Bryant and Dr. Kelly Maxwell who made this study so much more significant by sharing program quality data from earlier studies they completed in Georgia.
INTRODUCTION

Background of Georgia’s Compliance Determination System

Similar to other states, Georgia has a licensing and monitoring system that oversees a diverse population of early care and learning programs across the state. The licensing and monitoring system of early care and learning programs is charged to Bright from the Start: Georgia Department of Early Care and Learning (DECAL), a state early education department that also oversees and administers Georgia’s Pre-K Program, Child Care and Development Block Grant, the Child and Adult Care Food Program, and the Summer Food Service Program. In 2012, DECAL’s licensing and monitoring system regulated approximately 6,300 early care and learning programs. The crux of this regulation is determining if the programs meet Georgia’s health and safety rules. Programs that meet these rules are determined to be compliant.

In the mid 2000’s, Georgia began experimenting with a process that determined whether or not a program was designated as compliant with the state’s health and safety regulations by focusing on key Core Rules. These are health and safety rules deemed crucial to minimizing risk related to children’s health and safety. Seventy-four rules out of the 456 that programs must follow were classified as Core Rules\(^1\). Core Rules are cited by severity (low, medium, high, extreme). It is important to note that this entails a risk assessment theoretical approach rather than a Key Indicator statistical approach. This means that the Core Rules were determined by content analysis rather than by a statistical procedure.

Though this system has undergone some slight revisions, this basic methodology is still in place:

1. All programs receive at least one full licensing study and one monitoring visit. At the licensing study all applicable rules are examined. At the monitoring visit, only Core Rules (or any rule that was not met at the licensing study) are examined.
2. If additional visits are conducted, the Core Rules are examined again at that time.
3. At the end of the fiscal year (June 30), each program receives a compliance determination. This determination is based on all visits (licensing study, monitoring visit, and other reviews). A standardized worksheet, Annual Compliance Determination Worksheet (ACDW), is used to make the computation that determines the designation.
4. The compliance status remains until the next determination one year later. Programs do not have an opportunity to contest the compliance determination, though programs have numerous opportunities to contest any citation.
5. At the conclusion of Fiscal Year 2012, approximately 91% of the programs were classified as compliant. A program’s eligibility for certain services, acceptance into Quality Rated and Georgia’s Pre-K Program, is impacted by the program’s compliance determination.

Background of this Study

Since the compliance determination system has been used for several years, key policymakers at DECAL requested an external review to validate if the system was operating as intended. Are the Core Rules a sufficient subsample to measure a program’s overall regulation with the state’s health and safety regulations? Furthermore, does the compliance determination formula appropriately differentiate compliant programs from non-compliant programs? In other words, is the computation a viable way to make this designation? And finally, does compliance determination serve as a sufficient indicator for other aspects of quality not addressed in Georgia’s health and safety rules?

The purpose of this study was to validate the aforementioned compliance determination process. This validation process utilized a unique analytical model that compared licensing data with previous key indicator research and ascertained if the Core Rules are an indication of a program’s overall compliance with the total population of licensing rules. Second, additional statistical analyses examined if the mathematical formula used to compute compliance was an appropriate configuration of the data that differentiated between those programs that adequately met basic health and safety rules (compliant) and those that did not (non-compliant). Finally, licensing data were

\(^{1}\) The number of Core Rules was expanded in 2012 to include increased enforcement and sanctions regarding transportation. The new Core Rules were not part of this analysis.
compared to a representative sample of quality data collected as part of a different study to examine the correlation between compliance and quality (see a further explanation of the sample in the Limitations Section of this report).

Specifically, the study addressed the following research questions:

1. Do the Core Rules for child care centers (CCC) and family child care (FCC) homes serve as overall Key Indicators of compliance?
2. Does the Annual Compliance Determination Worksheet (ACDW) appropriately designate programs as compliant or non-compliant related to health and safety?
3. Are the Core Rules related to program quality?

The following definitions are used in the study:

**Core Rules** = the rules determined to be of greatest importance and place children at greatest risk if not complied with. This approach is defined in the licensing literature as a risk assessment approach. Core Rules cover 12 regulatory areas and 74 specific rules. The Core Rules were the focal point of this validation study and are addressed in the first approach to validation – Standards and the first research question.

**ACDW** = Annual Compliance Determination Worksheet, the compliance decision-making system based on the Core Rules that can be used to determine the number of visits made to programs. The ACDW was the secondary focal point of this validation study and is addressed in the second approach to validation – Measures and the second research question.

**Key Indicators** = a differential monitoring approach that uses only those rules that statistically predict overall compliance with all the rules. In other words, if a program is 100% in compliance with the Key Indicators, the program will also be in substantial to full compliance with all rules. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators, the program will also have other areas of non-compliance with all the rules. In this study, eight Key Indicators rules were identified for CCC and nine Key Indicators rules for FCC (See pages 15-16 for the specific indicators and additional detail about the methodology). These are in addition to the Core Rules.

**Rule Violations or Citations** = occurs when a program does not meet a specific rule and is cited as being out of compliance with that rule. These individual rule violations/citations are summed to come up with total violation/citation scores on the Core Rules and on the Licensing Studies.

**Differential Monitoring** = a relatively new approach to determining the number of licensing visits made to programs and to what rules are reviewed during these visits. Two measurement tools drive differential monitoring: one is a Weighted Risk Assessment, and the other is a Key Indicator checklist. Weighted Risk Assessments determine how often a program will be visited while Key Indicator checklists determine what rules will be reviewed in the program. Differential monitoring is a powerful approach when Risk Assessment is combined with Key Indicators because a program is reviewed by the most critical rules and the most predictive rules. See Figure 1 which presents a Logic Model & Algorithm for Differential Monitoring (*DMLMA©*) (Fiene, 2012).

**Licensing Study** = a comprehensive review of a program where all child care rules are reviewed.

**Monitoring Visit** = an abbreviated form of a visit and review in which only a select group (Core Rules) of child care rules are reviewed.

**Program Quality** = for the purposes of this study, quality was measured in child care centers by the *Early Childhood Environment Rating Scale-Revised (ECERS-R)*, *Infant Toddler Environment Rating Scale-Revised (ITERS-R)* and in family child care homes by the *Family Child Care Environment Rating Scale-Revised (FCCERS-R)*. The program quality measures were used as part of the third approach to validation – Outputs and the third research question.
Scoring for Licensing Variables/Data Collection Protocols:

**Licensing Study** = the total number of rule violations for a specific facility.

**Core Rules** = the total number of core rule violations.

**ACDW/Compliance Designation** = the annual compliance determination taken from the Annual Compliance Determination Worksheet. Compliant [C] was coded as “1” in the data base; Non-Compliant [NC] was coded as “0” in the data base.

**Key Indicators** = these were generated by a statistical methodology based upon the ability of the specific rule to predict full compliance with all the rules. Data from the Licensing Studies were used to make this determination of key indicator rule status.

METHODOLOGY AND ANALYTICAL FRAMEWORK

Licensing data over a four-year period (2008-2012) and matching program quality data from a two-year period (2007-2008) were used in this study. Specifically, data from 104 child care centers and 147 family child care homes were analyzed. Data from licensing studies (all rules) and monitoring visits (selected rules) were utilized. Program quality data were provided by researchers from the FPG Child Development Institute at the University of North Carolina at Chapel Hill (FPG), and the FPG research team matched these data points with the licensing data provided by DECAL. (See the following website for the specific reports - http://decal.ga.gov/BftS/ResearchStudyOfQuality.aspx). All the data were analyzed by the Research Institute for Key Indicators.

Two models were used to frame the analysis: a Validation Framework that uses four approaches (Zellman & Fiene, 2012) to validating quality rating and improvement systems (QRIS) being applied to licensing systems; and a *Differential Monitoring Logic Model and Algorithm (DMLMA©)* (Fiene, 2012) were employed to answer the three research questions for this Validation Study. The validation approaches are described below; the DMLMA© is described at the beginning of the Findings Section of this report.

**The first validation approach** deals with examining the validity of key underlying concepts by assessing if basic components and standards are the right ones by examining levels of empirical and expert support. For this study, this approach used Key Indicators to validate the Core Rules since Risk Assessment and Key Indicators are differential monitoring approaches. This answers the first research question.

**The second validation approach** deals with examining the measurement strategy and the psychometric properties of the measures used by assessing whether the verification process for each rule is yielding accurate results. Properties of the key rules can be measured through inter-rater reliability on observational measures, scoring of documentation, and inter-item correlations to determine if measures are psychometrically sound. Cut scores can be examined to determine the most appropriate ways to combine measures into summary ratings. For this study, the second validation approach validates the use of the ACDW and Core Rules by comparing compliance decisions with the Licensing Studies. This answers the second research question.

**The third validation approach** deals with assessing the outputs of the licensing process by examining the variation and patterns of program level ratings within and across program types to ensure that the ratings are functioning as intended. The approach examines the relationship of program level ratings to other more broadly based program quality measures and examines alternate cut points and rules to determine how well the ratings distinguish different levels of quality. For this study, this approach used data from Core Rules and Licensing Studies and data from earlier program quality studies (Maxwell, et al., 2009a,b; 2010) for validation. This answers the third research question.

Out of the four validation approaches (See Table 8), only three were utilized in this study. **The fourth validation approach** deals with how ratings are associated with children’s outcomes. This approach examines the relationship...
between program level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes. This approach did not have data that could be used in this study.

FINDINGS

The DMLMA© (See Figure 1) provides the conceptual model for assessing the overall effectiveness of Georgia’s approach using Core Rules. In the model, the two main tools are Risk Assessment and Key Indicator measurements, which are created from a statistical analysis of the comprehensive licensing tool. The comprehensive licensing tool measures compliance with all rules. For the purposes of this study the Licensing Study represents the comprehensive licensing tool while the Core Rules represent a Risk Assessment tool. For the Program Quality tools, the ECERS-R, ITERS-R and FCCERS-R were utilized from an earlier program quality study by FPG Child Development Institute at the University of North Carolina at Chapel Hill (Maxwell, et al., 2009a,b; 2010). Georgia currently does not use a Key Indicator tool (see Table 1). With the DMLMA© analytical methodology, specific correlational thresholds are expected (please refer to Figure 1 on page 14).

### TABLE 1

<table>
<thead>
<tr>
<th>DMLMA© Terminology</th>
<th>Georgia Examples and Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive Tool</td>
<td>Licensing Study</td>
</tr>
<tr>
<td>Program Quality Tool</td>
<td>ECERS-R and ITERS-R for CCC; FCCERS-R for FCC</td>
</tr>
<tr>
<td>Risk Assessment Tool</td>
<td>Core Rules</td>
</tr>
<tr>
<td>Key Indicators Tool</td>
<td>Not Present (Generated as part of this Study-see Tables 9/10)</td>
</tr>
<tr>
<td>Differential Monitoring Tool</td>
<td>ACDW Compliance Determination</td>
</tr>
</tbody>
</table>

Before presenting the findings for the validation approaches, some basic descriptive statistics are provided regarding the major variables in this study: Licensing Study, ACDW, Core Rules, and Key Indicators (see Table 2). The data are provided for both child care centers and family child care homes. It is clear from these basic descriptive statistics that the data distributions are very skewed in a positive fashion which means that there is very high compliance with all the major licensing variables for this study. In other words, the majority of programs are in substantial compliance with all the licensing rules and receive a compliant determination.

### TABLE 2

<table>
<thead>
<tr>
<th>Licensing Variable</th>
<th>Mean</th>
<th>Range</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Licensing Study (CCC)</td>
<td>5.51</td>
<td>25</td>
<td>5.26</td>
<td>1.47</td>
<td>2.11</td>
</tr>
<tr>
<td>ACDW (CCC)</td>
<td>0.75</td>
<td>1</td>
<td>0.44</td>
<td>-1.17</td>
<td>-0.64</td>
</tr>
<tr>
<td>Core Rules (CCC)</td>
<td>4.47</td>
<td>22</td>
<td>4.72</td>
<td>1.81</td>
<td>3.60</td>
</tr>
<tr>
<td>Key Indicators (CCC)</td>
<td>1.68</td>
<td>6</td>
<td>1.61</td>
<td>0.90</td>
<td>0.073</td>
</tr>
<tr>
<td>Licensing Study (FCC)</td>
<td>5.85</td>
<td>33</td>
<td>5.71</td>
<td>1.56</td>
<td>3.37</td>
</tr>
<tr>
<td>ACDW (FCC)</td>
<td>0.87</td>
<td>1</td>
<td>0.34</td>
<td>-2.23</td>
<td>3.03</td>
</tr>
<tr>
<td>Core Rules (FCC)</td>
<td>1.61</td>
<td>11</td>
<td>1.75</td>
<td>1.99</td>
<td>6.61</td>
</tr>
<tr>
<td>Key Indicators (FCC)</td>
<td>2.37</td>
<td>8</td>
<td>2.13</td>
<td>0.63</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

Licensing Study Mean = the average number of total rule violations.  
ACDW Mean = the average score for a determination of compliance (1) or non-compliance (0).  
Core Rules Mean = the average number of core rule violations.  
Key Indicators Mean = the average number of key indicator violations.

The findings are presented by the three validation approaches of Standards, Measures, and Outputs as well as the three research questions related to Key Indicators, Core Rules, and Program Quality.

1) Validation of Standards (First Approach to Validation) for answering the first research question: Do the Core Rules for child care centers (CCC) and family child care (FCC) homes serve as overall key indicators of compliance?

In this first approach to validation which focuses on Standards, Key Indicators were generated from the Licensing Studies because Core Rules (a Risk Assessment tool) and Key Indicators are both Differential Monitoring approaches (see Figure 1). The Core Rules were compared to the Key Indicators generated by the licensing data base and there was a .49 correlation for CCC (n = 104) and .57 correlation for FCC (n = 147) which indicates a
relationship between the Core Rules and Key Indicators at a p < .0001 significance level (Table 3). Also, the Key Indicators were correlated with the Licensing Study data and significant results were determined with r values of .78 (p < .0001) for CCC (n = 104) and .87 (p < .0001) for FCC (n = 147). These results clearly met the expected DMLMA© thresholds between the key indicator rules with core rules (.50+) and licensing studies (.70+).

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Key Indicators with Core Rules and Licensing Study</th>
<th>r</th>
<th>p</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Key Indicators and Core Rules (CCC)</td>
<td>.49</td>
<td>.0001</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Key Indicators and Licensing Study (CCC)</td>
<td>.78</td>
<td>.0001</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Key Indicators and Core Rules (FCC)</td>
<td>.57</td>
<td>.0001</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Key Indicators and Licensing Study (FCC)</td>
<td>.87</td>
<td>.0001</td>
<td>147</td>
</tr>
</tbody>
</table>

Table 3 begins to demonstrate how the Georgia Child Care Licensing system is utilizing the DMLMA© terminology from Table 1. With the generation of Key Indicators from this study, all the key elements within a differential monitoring system are present. This crosswalk to the DMLMA© will continue in Tables 4 & 5.

2) Validation of Measures (Second Approach to Validation) for answering the second research question: Is the Annual Compliance Determination Worksheet (ACDW) a valid measure in determining the overall health and safety compliance of Georgia’s early care and learning programs?

The Core Rules and the ACDW were compared to the Licensing Study data and compliance designation to determine the validation of the ACDW scoring protocol. There was a high correlation between the number of violations on the Core Rules and the total licensing violations on the Licensing Studies (r = .69; p < .0001) (Table 4). This result helps to validate that the ACDW is actually discriminating between high compliant and low compliant providers for CCC. For FCC, there was also a high correlation between the number of violations on the Core Rules and the total licensing violations on the Licensing Studies (r = .74; p < .0001). These results meet the DMLMA© thresholds of .50+ for Licensing Studies and Core Rules.

When Core Rules were correlated with the ACDW compliance decisions, there was a significantly high correlation for CCC (r = .76; p < .0001) and for FCC (r = .70; p < .0001). The key element of the ACDW scoring protocol is that the Core Rules distinguish between high and low compliant providers. The CCC/Core Rules and ACDW have been validated, as well as the FCC/Core Rules and ACDW because both the correlations were above the expected DMLMA© threshold (.50+).

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Core Rules with Licensing Studies and ACDW</th>
<th>r</th>
<th>p</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Core Rules and Licensing Studies (CCC)</td>
<td>.69</td>
<td>.0001</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Core Rules and ACDW (CCC)</td>
<td>.76</td>
<td>.0001</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Core Rules and Licensing Studies (FCC)</td>
<td>.74</td>
<td>.0001</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Core Rules and ACDW (FCC)</td>
<td>.70</td>
<td>.0001</td>
<td>147</td>
</tr>
</tbody>
</table>

3) Validation of Outputs (Third Approach to Validation) for answering the third research question: Are the Core Rules correlated with program quality?

For this approach, programs were divided into those that had an ITERS-R score, an ECERS-R score for a preschool class, and an ECERS-R score for a Georgia’s Pre-K class; and those that had only an ITERS-R score and an ECERS-R score for preschool. The sample was evenly divided. Since Georgia has placed substantial resources into its Pre-K program, it was thought that this analysis might suggest if there was anything different between programs with a Georgia’s Pre-K class and those without.

When the Core Rules for CCC’s were compared with program quality data (ECERS-R/PS + ITERS-R), a significant correlation was not found between CCC (r = .27) for programs with only preschool classrooms but was found for programs with Pre-K classrooms (ECERS-R/PK + ITERS-R) (r = .60). When Core Rules for FCC’s were compared
to the FCC program quality data (FCCERS-R), the correlations were at a much lower level (r = .17) (See Table 5). However, these results are constrained by the limited range of the data; see the Limitation Section that follows this section.

Upon closer inspection of the correlations in Table 5 for CCC, it would appear that the CCC compliance system is more valid with the state-funded Pre-K programs (.48) than with the preschool programs (.21) because the correlations between the various Environment Rating Scales (ECERS-R + ITERS-R) are significant only when compared to the respective compliance with all rules on the Licensing Studies in the programs that have Pre-K programs. In making these comparisons, programs that had both ECERS-R and ITERS-R were combined and compared to the respective Licensing Study data (these data were reversed scored in which the number of violations were subtracted from a perfect score of 100). The differences are even more significant when you compare the Environment Rating Scales and the Core Rules where the Pre-K programs’ correlation between the compliance with Core Rules and Environment Rating Scales is .60 and preschool programs is .27 while the FCC is .17.

Program quality data refer to data collected in earlier studies by researchers from FPG (Maxwell, et al., 2009a,b; 2010) in which FPG collected Environment Rating Scales (ECERS-R; ITERS-R; FCCERS-R) data on a representative sample of CCC and FCC. In comparing the program compliance and program quality data, the analyses supported the validation of the CCC for Pre-K only programs (DMLMA© threshold = .30+) but it was weaker for the FCC programs and not significant for preschool programs and therefore could not be validated. See Table 13 on page 17 for a further explanation of the CCC data distribution.

![Table 5](http://decal.ga.gov/BfjS/ResearchStudyOfQuality.aspx)

**TABLE 5**

<table>
<thead>
<tr>
<th>Program Compliance and Quality Comparisons</th>
<th>r</th>
<th>p</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECERS-R/PK + ITERS-R and Licensing Studies</td>
<td>.48</td>
<td>.001</td>
<td>45</td>
</tr>
<tr>
<td>ECERS-R/PK + ITERS-R and Core Rules</td>
<td>.60</td>
<td>.0001</td>
<td>45</td>
</tr>
<tr>
<td>ECERS-R/PS + ITERS-R and Licensing Studies</td>
<td>.21</td>
<td>ns</td>
<td>45</td>
</tr>
<tr>
<td>ECERS-R/PS + ITERS-Rand Core Rules</td>
<td>.27</td>
<td>ns</td>
<td>45</td>
</tr>
<tr>
<td>FCCERS-R and Licensing Studies</td>
<td>.19</td>
<td>.04</td>
<td>146</td>
</tr>
<tr>
<td>FCCERS-R and Core Rules</td>
<td>.17</td>
<td>.03</td>
<td>146</td>
</tr>
</tbody>
</table>

**LIMITATION**

The sampling for this study was based on previous studies (Maxwell, 2009a,b; 2010) completed by FPG in which program quality data were collected and analyzed. This study employed a subset of sites that were a representative sample of Georgia’s child care licensing system. Not all of these sites could be used for this study because some had closed or some did not have the necessary data to make comparisons. So the sample at this point is one of convenience; however, 104 of the 173 CCC and 146 of the 155 FCC were used in this study, a significant number of the original representative sample. Also, when the Environment Rating Scales (ECERS-R, ITERS-R, FCCERS-R) scores were compared with the CCC and FCC samples, there were no significant differences (average difference was .01-.03) between the two study samples (See Table 6).

![Table 6](http://decal.ga.gov/BfjS/ResearchStudyOfQuality.aspx)

**TABLE 6**

<table>
<thead>
<tr>
<th>Environment Rating Scale Scores</th>
<th>FPG</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECERS-R Pre-K Total Scale Scores</td>
<td>4.16</td>
<td>4.15</td>
</tr>
<tr>
<td>ECERS-R Preschool Total Scale Scores</td>
<td>3.39</td>
<td>3.42</td>
</tr>
<tr>
<td>ITERS-R Total Scale Scores</td>
<td>2.74</td>
<td>2.72</td>
</tr>
<tr>
<td>FCCERS-R Total Scale Scores</td>
<td>2.50</td>
<td>2.49</td>
</tr>
</tbody>
</table>
CONCLUSION

The CCC differential monitoring through the Core Rules/ACDW has been validated on the three approaches (Standards, Measures, and Outputs (Pre-K Program only)) and three research questions (Key Indicators, Core Rules, Program Quality (Programs with Georgia Pre-K only)) (See Table 7). The FCC differential monitoring through the Core Rules/ACDW was validated on the first validation approach (Standards) and first research question (Key Indicators); validated on the second validation approach (Measures) and second research question (Core Rules); but not validated on the third validation approach (Outputs) and third research question (Program Quality).

TABLE 7

<table>
<thead>
<tr>
<th>Validation Approach/Research Question</th>
<th>CCC Actual (Expected*)</th>
<th>FCC Actual (Expected)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 STANDARDS/Key Indicators</td>
<td>VALIDATED</td>
<td>VALIDATED</td>
</tr>
<tr>
<td>Key Indicators x Core Rules</td>
<td>.49 (.50+)</td>
<td>.57 (.50+)</td>
</tr>
<tr>
<td>Key Indicators x Licensing Studies</td>
<td>.78 (.70+)</td>
<td>.87 (.70+)</td>
</tr>
<tr>
<td>2 MEASURES/Core Rules/ACDW²</td>
<td>VALIDATED</td>
<td>VALIDATED</td>
</tr>
<tr>
<td>Core Rules x Licensing Studies</td>
<td>.69 (.50+)</td>
<td>.74 (.50+)</td>
</tr>
<tr>
<td>Core Rules x ACDW</td>
<td>.76 (.50+)</td>
<td>.70 (.50+)</td>
</tr>
<tr>
<td>3 OUTPUTS/Program Quality</td>
<td>VALIDATED</td>
<td>NOT VALIDATED</td>
</tr>
<tr>
<td>Licensing Studies x ERS**/PK</td>
<td>.48 (.30+)</td>
<td>FCCERS .19 (.30+)</td>
</tr>
<tr>
<td>Core Rules x ERS/PK</td>
<td>.60 (.30+)</td>
<td>FCCERS .17 (.30+)</td>
</tr>
<tr>
<td>Licensing Studies x ERS/PS</td>
<td>---------</td>
<td>.21 (.30+)</td>
</tr>
<tr>
<td>Core Rules x ERS/PS</td>
<td>---------</td>
<td>.27 (.30+)</td>
</tr>
</tbody>
</table>

*Expected r Value Thresholds in Order to be Validated (Also see Figure 1 for additional details):
High correlations (.70+) = Licensing Studies x Key Indicators; Core Rules x Licensing Studies; ERS x Core Rules; Core Rules x ACDW; Core Rules x Key Indicators; Key Indicators x ACDW. Moderate correlations (.50+) = Program Quality Tools x Licensing Studies; Program Quality x Core Rules; Program Quality x Key Indicators. Lower correlations (.30+) = Pre-K program x Licensing Studies; Pre-K program x Core Rules; Pre-K program x Key Indicators.

Program Quality Tools = ECERS-R, ITERS-R, FCCERS-R.

**ERS = ECERS-R + ITERS-R
PK = Pre-K program
PS = Preschool program

A confounding of data occurred with the first two validation approaches because the Core Rules were influenced great deal by the National Child Care Key Indicators (NCCKI) (Fiene, 2002) where 10 of the 13 Core Rules overlapped significantly with the NCCKI. This helped to increase the correlation between the Core Rules and the Licensing Studies because the Core Rules represented both risk assessment and key indicator rules. Using both risk assessment and key indicator rules together is an ideal differential monitoring approach (Fiene, 2012). Most states use one or the other but generally not together. By including the newly generated key indicators from this study where there is also overlap with the NCCKI, it should enhance the differential monitoring approach utilized by DÉCAL.

². ACDW decisions were compared with using severity as a factor and not using it as a factor in the scoring system with Core Rules. No significant differences were found between the two scoring systems; therefore, the results in this study represent Core Rule scores without severity included since this is the simpler model.
RECOMMENDATIONS

The following recommendations can be made from this Licensing Differential Monitoring Validation Study.

1) **First research question/validation recommendation:** Revise the worksheet determination scoring relative to the visiting protocol by combining the Core Rules with a Key Indicator approach so that if any of the Core Rules or Key Indicators are out of compliance, then a full compliance review (Licensing Study) should be used. The present worksheet determination scoring protocol is overly complex. Just moving to a more comprehensive review (Licensing Study) based on non-compliance with the Core Rules will simplify the scoring protocol and make determinations more straightforward. If there is full (100%) compliance with the Core Rules and Key Indicators, then the next scheduled review of the program would be an abbreviated Monitoring Visit. If there is not 100% compliance with the Core Rules and Key Indicators, then the next scheduled review of the program would be a Licensing Study reviewing all child care rules. Based upon the compliance/non-compliance scores of the Licensing Study will determine how often the program will be visited. A revised Georgia Differential Monitoring System could potentially look like the following:

![Diagram of Compliance Decisions]

**Compliance Decisions:**

- **Core Indicators Screener = Core Rules + Key Indicators**
  - This becomes a screening tool to determine if a program receives a Licensing Study reviewing all child care rules or an abbreviated Monitoring visit continuing to review key indicator and core rules for their next visit.

- **Core Indicators (100%)** = the next visit is a Monitoring Visit. Every 3-4 years a full Licensing Study is conducted.

- **Core Indicators (not 100%)** = The next visit is a Licensing Study where all rules are reviewed.

- **Compliance** = 96%+ with all rules and 100% with Core Indicators. The next visit is a Monitoring Visit.

- **Non-compliance** = less than 96% with all rules. The next visit is a Licensing Study.

2) **Second research question/validation recommendation:** Follow the development of weighted risk assessment tools as outlined by Fiene & Kroh (2000) in the NARA Licensing Chapter for CCC and FCC. It has been over 20 years since Core Rules were weighted. It is recommended that Core Rules be weighted every 10 years. Doing a weighted risk assessment would help confirm that the present Core Rules are the highest risk rules.

3) **Third research question/validation recommendation:** Confirm the CCC (ERS/PS) and FCC results by conducting a more recent program quality study that reflects all the changes made within the CCC and FCC systems. Although FCC program quality and Licensing Study and Core Rules reached statistical significance, the overall correlation was too low (Licensing Studies = .19; Core Rules = .17). With the CCC system the Pre-K program demonstrated significant correlations between ERS/PK and Licensing Study (.48) & Core Rules (.60) but not the Preschool program (ERS/PS: Licensing Studies = .21; Core Rules = .27).

---

These recommendations are drawn from the data in this study and previous studies conducted by the author in which the empirical evidence led to similar recommendations.
REFERENCES


<table>
<thead>
<tr>
<th>Approach</th>
<th>Activities and Purpose</th>
<th>Typical Questions Approach Addresses</th>
<th>Issues and Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Examine the validity of key underlying concepts</strong></td>
<td>Assess whether basic QRIS quality components and standards are the “right” ones by examining levels of empirical and expert support.</td>
<td>Do the quality components capture the key elements of quality? Is there sufficient empirical and expert support for including each standard?</td>
<td>Different QRISs may use different decision rules about what standards to include in the system.</td>
</tr>
<tr>
<td><strong>2. Examine the measurement strategy and the psychometric properties of the measures used to assess quality</strong></td>
<td>Examine whether the process used to document and verify each indicator is yielding accurate results. 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 psychometrically sound. Examine the relationships among the component measures to assess whether they are functioning as expected. Examine cut scores and combining rules to determine the most appropriate ways to combine measures of quality standards into summary ratings.</td>
<td>What is the reliability and accuracy of indicators assessed through program administrator self-report or by document review? What is the reliability and accuracy of indicators assessed through observation? Do quality measures perform as expected? (e.g., do subscales emerge as intended by the authors of the measures?) Do measures of similar standards relate more closely to each other than to other measures? Do measures relate to each other in ways consistent with theory? Do different cut scores produce better rating distributions (e.g., programs across all levels rather than programs at only one or two levels) or more meaningful distinctions among programs?</td>
<td>This validation activity is especially important given that some component measures were likely developed in low-stakes settings and have not been examined in the context of QRIS.</td>
</tr>
</tbody>
</table>
TABLE 8 (CONTINUED)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Activities and Purpose</th>
<th>Typical Questions Approach Addresses</th>
<th>Issues and Limitations</th>
</tr>
</thead>
<tbody>
<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 ensure that the ratings are functioning as intended. Examine relationship of program-level ratings to other quality indicators to determine if ratings are assessing quality in expected ways. Examine alternate cut points and rules to determine how well the ratings distinguish different levels of quality.</td>
<td>Do programs with different program-level ratings differ in meaningful ways on alternative quality measures? Do rating distributions vary by program type, e.g., ratings of center-based programs compared to ratings of home-based programs? Are current cut scores and combining rules producing appropriate distributions across rating levels?</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>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 children who attend higher-rated programs have greater gains in skills than children who attend lower-quality programs?</td>
<td>Appropriate demographic and program level control variables must be included in analyses to account for selection factors. Studies could be done on child and program samples to save resources. Findings do not permit attribution of causality about QRIS participation but inferences can be made about how quality influences children’s outcomes.</td>
</tr>
</tbody>
</table>
**FIGURE 1- DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (Fiene, 2012)**

*DMLMA© Applied to the Georgia Child Care Licensing System*

CI + PQ => RA + KI => DM

**Georgia Examples:**

CI = Comprehensive Tool = Licensing Study (LS – All Rules)
PQ = Program Quality Tool = Environmental Rating Scales (ERS = ECERS-R, ITERS-R, FCCERS-R)
RA = Risk Assessment Tool = Core Rules (CR)
KI = Key Indicators Tool = presently Georgia does not have a KI
DM = Differential Monitoring Tool = ACDW (Compliance/Non-Compliance Decision)

A very important concept in this validation study is that the system employed by DECAL is a risk assessment approach rather than a key indicator methodology which is based upon predictor rules. The *DMLMA©* is a new methodology assessing the effectiveness and efficiency of Differential Monitoring systems being used by state regulatory agencies and provides the conceptual model for this study.

---

**DMLMA© Thresholds:**

*High Correlations (0.70+) = CI x KI.*

*Moderate Correlations (0.50+) = CI x RA; RA x DM; RA x KI; KI x DM.*

*Lower Correlations (0.30+) = PQ x CI; PQ x RA; PQ x KI.*
### Table 9 - Listing of Key Indicators for Georgia Child Care Centers with Phi Coefficients

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description</th>
<th>Phi</th>
</tr>
</thead>
<tbody>
<tr>
<td>591-1-1-25 (3)</td>
<td>Requires that the center and surrounding premises be clean, free of debris and in good repair.</td>
<td>.49</td>
</tr>
<tr>
<td>591-1-1-.25 (13)</td>
<td>Requires that hazardous equipment, materials and supplies be inaccessible to children.</td>
<td>.46</td>
</tr>
<tr>
<td>591-1-1-.26 (6)</td>
<td>Requires that outdoor equipment be free of hazards such as lead-based paint, sharp corners, rust and splinters.</td>
<td>.44</td>
</tr>
<tr>
<td>591-1-1-.26 (8)</td>
<td>Requires the playground to be kept clean, free of litter and hazards.</td>
<td>.59</td>
</tr>
<tr>
<td>591-1-1.26 (7)</td>
<td>Requires that a resilient surface be provided and maintained beneath the fall zone of climbing and swinging equipment.</td>
<td>.57</td>
</tr>
<tr>
<td>591-1-1-.36 (6)(a-c)</td>
<td>Requires the center to maintain on the vehicle current information for each child including a) center and passenger information; b) emergency medical information and c) a passenger checklist.</td>
<td>.49</td>
</tr>
<tr>
<td>591-1-1-.14 (1)</td>
<td>Requires that at least 50% of the caregiver staff have current first aid and CPR training.</td>
<td>.49</td>
</tr>
<tr>
<td>591-1-1-.08 (a)-(f)</td>
<td>Requires the center to maintain a file for each child while such child is in care and for one year after that child is no longer enrolled.</td>
<td>.44</td>
</tr>
</tbody>
</table>

### Table 10 - Listing of Key Indicators for Georgia Family Child Care Homes with Phi Coefficients

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description</th>
<th>Phi</th>
</tr>
</thead>
<tbody>
<tr>
<td>290-2.3-.11(2)(C)</td>
<td>Requires that fire drills be practiced monthly and shall be documented and kept on file for one year.</td>
<td>.51</td>
</tr>
<tr>
<td>290-2-3-.11 (2)(f)</td>
<td>Requires that poisons, medicines, cleaning agents and other hazardous materials be in locked areas or inaccessible to children.</td>
<td>.61</td>
</tr>
<tr>
<td>290-2-3-.11 (1)(f)</td>
<td>Requires the family day care home and any vehicle used to have a first aid kit.</td>
<td>.57</td>
</tr>
<tr>
<td>290-2-3-.07 (4)</td>
<td>Requires that the provider obtain ten clock hours of training in child care issues from an approved source within the first year and thereafter on an annual basis.</td>
<td>.58</td>
</tr>
<tr>
<td>290-2-3-.08 (1)(a)</td>
<td>Requires the family day care home to maintain a file for each child that includes the child’s name, birth date, parents or guardian’s name, home and business addresses and telephone numbers.</td>
<td>.63</td>
</tr>
<tr>
<td>290-2-3-.08 (1)(b)</td>
<td>Requires that the record for each child contain the names(s), address(es) and telephone number(s) of person(s) to contact in emergencies when the parent cannot be reached.</td>
<td>.57</td>
</tr>
<tr>
<td>290-2-3-.08 (1)(b)</td>
<td>Requires the family day care home to maintain a file for each child that includes the name, address and telephone number of the child’s physician to contact in emergencies.</td>
<td>.55</td>
</tr>
<tr>
<td>290-2-3-.08 (1)(f)</td>
<td>Requires the family day care home to maintain a file for each child that includes known allergies, physical problems, mental health disorders, mental retardation or developmental disabilities which would limit the child’s participation in the program.</td>
<td>.51</td>
</tr>
<tr>
<td>290-2-3-.08 (1)(c)</td>
<td>Requires the family day care home to maintain a file for each child that includes evidence of age appropriate immunizations or a signed affidavit against such immunizations; enrollment in the home may not continue for more than 30 days without such evidence.</td>
<td>.72</td>
</tr>
</tbody>
</table>
Table 11 - Key Indicator Formula Matrix for Generating Key Indicators*

<table>
<thead>
<tr>
<th></th>
<th>Providers In Compliance on Rule</th>
<th>Programs Out Of Compliance on Rule</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Group</strong></td>
<td>A</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Low Group</strong></td>
<td>C</td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td><strong>Column Total</strong></td>
<td>W</td>
<td>X</td>
<td>Grand Total</td>
</tr>
</tbody>
</table>

(* This computation occurred for each licensing rule)

Figure 2 - Key Indicator Statistical Methodology (Calculating the Phi Coefficient)

\[
\phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}}
\]

\[A = \text{High Group} + \text{Programs in Compliance on Specific Rule.}\]
\[B = \text{High Group} + \text{Programs out of Compliance on Specific Rule.}\]
\[C = \text{Low Group} + \text{Programs in Compliance on Specific Rule.}\]
\[D = \text{Low Group} + \text{Programs out of Compliance on Specific Rule.}\]
\[W = \text{Total Number of Programs in Compliance on Specific Rule.}\]
\[X = \text{Total Number of Programs out of Compliance on Specific Rule.}\]
\[Y = \text{Total Number of Programs in High Group.}\]
\[Z = \text{Total Number of Programs in Low Group}\]

**High Group** = Top 25% of Programs in Compliance with all Rules.
**Low Group** = Bottom 25% of Programs in Compliance with all Rules.

Table 12 – Phi Coefficient Decision Table

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include</td>
</tr>
<tr>
<td>(+.25) – (-.25)</td>
<td>Unpredictable</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>
Table 13 - Comparison of the Pre-K and Preschool Programs

<table>
<thead>
<tr>
<th>Compliance Level*</th>
<th>Pre-K ECERS-R**(N)</th>
<th>Preschool ECERS-R***((N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>4.88 (4)</td>
<td>3.40 (15)</td>
</tr>
<tr>
<td>99</td>
<td>4.13 (6)</td>
<td>4.35 (7)</td>
</tr>
<tr>
<td>98</td>
<td>4.38 (6)</td>
<td>3.89 (13)</td>
</tr>
<tr>
<td>97</td>
<td>3.99 (4)</td>
<td>3.15 (9)</td>
</tr>
<tr>
<td>96</td>
<td>4.36 (2)</td>
<td>3.16 (13)</td>
</tr>
<tr>
<td>95</td>
<td>4.60 (2)</td>
<td>3.53 (5)</td>
</tr>
<tr>
<td>90</td>
<td>3.43 (2)</td>
<td>2.56 (5)</td>
</tr>
<tr>
<td>80</td>
<td>2.56 (1)</td>
<td>2.38 (2)</td>
</tr>
</tbody>
</table>

*Compliance Level = the number of child care rule violations subtracted from 100.

100 = Full Compliance with Rules
99-98 = Substantial Compliance with Rules
97-90 = Medium Level of Compliance with Rules
80 = Low Level of Compliance with Rules

**Pre-K ECERS-R = average score of Pre-K Program classrooms as compared to the respective compliance levels. (N) = Sample Size.

***Preschool ECERS-R = average score of Preschool Program classrooms as compared to the respective compliance levels. (N) = Sample Size.

From this comparison there is more of a linear relationship between compliance levels and ECERS-R average scores for Pre-K Program classrooms than with the Preschool Program classrooms where there is more of a curvilinear or plateau effect at the upper end of compliance levels (Full Compliance). In order to attain the necessary correlational thresholds (+.30+) for validation for the third approach to validation, having a linear relationship rather than curvilinear will enhance this occurring. When a curvilinear or plateau effect occurs there is too great a likelihood that programs at a medium level of quality will be introduced into the highest (full) level of compliance. From a public policy standpoint this is an undesirable result.

The other item to note with the data distributions is that the Preschool ECERS-R data are more restricted than the Pre-K Program ECERS-R data. In other words, there is less variance in the Preschool Program ECERS-R data than in the Pre-K Program ECERS-R data.

There is an important limitation in these data that the reader must be aware of in not drawing any conclusions that the presence of a Pre-K Program classroom in any way is causing the change in licensing compliance. There is a relationship between the two but there is no assumption of causality.
Because of the nature of this report being a state’s first attempt at fully validating it’s Child Care Licensing Core Rule Differential Monitoring Approach utilizing the Zellman & Fiene (2012) Validation Framework and Fiene’s DMLMA (2012) Model, certain questions surfaced regarding the terminology and the methodology being used in this report. This Technical Elements Appendix provides answers to specific questions that have been raised regarding these methodologies.

1. **How were the multiple years of data handled?**

   The Licensing Study data used to make the comparisons are the facility reports that were the earliest facility observations so that these data would be closest to when the program quality data were collected. The other more recent Licensing Studies were not used in this comparison.

2. **If the Core Rules, Key Indicator, and Licensing Study values are counts of violations, how was the fact that different sites had different numbers of visits handled?**

   Because only the earliest Licensing Study data was used, the number of visits were not an issue in the scoring.

3. **If the Core Rules, Key Indicator, and Licensing Study values are counts of violations, were all levels of violation risk (low, medium, high, extreme) handled the same?**

   Yes, there were very few occurrences of high and extreme in the data base and also no significant differences were found when a sample of the rule violations with and without the levels of violation risk were compared. Therefore the simpler formula in which levels of violation risk were not used was selected.

4. **How did you determine the minimum correlations (DMLMA thresholds) for each analysis? Was this computed separately for this analysis or are the minimum correlations based on previous work?**

   The DMLMA thresholds were determined from previous research work conducted by the author of this study on this model over the past 30 years. These were the average correlational thresholds that have been proposed for making validation determinations. The reason for utilizing the DMLMA model and thresholds is that the Zellman & Fiene (2012) Framework provides guidance in how to select specific validation approaches, what are the specific questions answered by the approach and what are the limitations of the particular approach. The DMLMA model builds upon this but provides a suggested scoring protocol by comparing correlational thresholds in a specific state to historical trends.

5. **Was Phi calculated for every rule in the licensing study? Can the full list be added to the appendix?**

   Yes, Phi was calculated for every rule in the licensing study but most of them could not be computed because there was so few rule violations in the majority of the rules. This is typical of state licensing data sets and the full Phi comparisons are not depicted because it does not add any information to the state report.
6. **How did you determine which of the Licensing Study rules should be counted as Key Indicators?**

The Key Indicator statistical methodology based upon a specific cut off point for the Phi Coefficient in which the p values were .0001 or less. This is a very stringent cut off point but it has been found historically that the p values needed to be lowered as the data distributions became more skewed with programs overall compliance levels increasing over time.

7. **How were sites that had no infant/toddler (i.e., no ITERS score) handled for the third validation approach? How were sites that had only a GA Pre-K (no preschool) handled?**

For scoring purposes only those facilities that had both the ECERS and ITERS scores were used in making comparisons with the licensing data related to the third approach to validation. The GA Pre-K were scored and compared in the same way.

8. **On Table 13, why is the number of violation subtracted from 100 (rather than from the maximum possible)?**

Generally this scoring is done because it is more intuitive to think in terms of 100% in compliance as a score of “100” rather than a score of “0”. This conversion is used in all state licensing reports that involve the DMLMA, Key Indicators and Risk Assessment Models.
FOR ADDITIONAL INFORMATION REGARDING THIS REPORT AND STUDY:

Bentley Ponder, Ph.D.
DECAL Director of Research and Evaluation
Bentley.Ponder@decal.ga.gov

Richard Fiene, Ph.D.
Research Institute for Key Indicators LLC
DrFiene@gmail.com
ABSTRACT

This report will provide a blueprint for Oregon’s Early Care and Education/Child Care program monitoring system in developing a Differential Program Monitoring, Risk Assessment, and Key Indicator approach to help streamline their present licensing process. The report will be organized into the following major headings: an introduction to the differential monitoring methodology; how key indicators and risk assessment fit into the larger program monitoring of early care and education programs; how key indicators and risk assessment will be applied to Oregon’s system in particular; the technical aspects of differential monitoring, risk assessment and key indicator methodology, the sample to be drawn from the population, potential results from the analyses; a timeline for this developmental effort; and potential cost savings from the approach.

INTRODUCTION

The Risk Assessment, Key Indicator, and Differential Program Monitoring Methodologies were developed to help streamline the program monitoring of early care and education programs. It was first applied in child care licensing (Fiene & Nixon, 1985) but has been used in many other service types, such as: Head Start Performance Standards (Fiene, 2013a), National Accreditation (Fiene, 1996), and child and adult residential programs (Kroh & Melusky, 2010). The methodologies are based upon statistical protocols that have been developed in the tests and measurements literature in which an abbreviated set of items is used to statistically predict as if the full test was applied. This methodology has been used in regulatory analysis and more recently has been proposed for use in Quality Rating and Improvement Systems (QRIS) (Fiene, 2013b).
DIFFERENTIAL PROGRAM MONITORING

Risk Assessment and Key Indicators are important components of differential program monitoring which employs an abbreviated review rather than a comprehensive or full review of a program. It is one of several key elements that have been identified in the research literature to help improve the cost effectiveness and efficiency of the program monitoring of early care and education programs (Fiene, 2013b, c) (See the Appendix). A recent addition to differential monitoring are QRIS – Quality Rating and Improvement Systems. Key indicators have a long history of development within the licensing literature (Fiene & Kroh, 2000) but have only recently been proposed to be used with QRIS. This proposed blueprint is to assist Oregon to develop a fully functional differential program monitoring, risk assessment, and key indicator approach to their child care licensing system and then determine the feasibility of using these approaches with its QRIS system.

The other key elements of the differential program monitoring approach are the following: program compliance/licensing which is generally a state’s health and safety rules/regulations that govern child care. At the national level this would be Caring for Our Children: National Performance Standards for Health and Safety in Child Care (2012). The program quality key element is generally represented by the state’s QRIS. At the national level it is represented by accreditation, such as NAEYC, NECPA, or NAFCC. The key indicator element is represented by the state’s statistical predictor rules/regulations drawn from their comprehensive set of health and safety rules/regulations that govern child care. At the national level, an example is the 13 Indicator of Quality Child Care (2002). This element can also represent a state’s statistical predictor QRIS standards drawn from the comprehensive set of QRIS standards. The purpose of this Blueprint Report is to develop these statistically predictor standards first for Oregon’s child care licensing system and explore the possibility of expanding this to their QRIS system. The last key element to be addressed in this report is the risk assessment key element in which these are the high risk rules/regulations that place children at greatest risk of mortality or morbidity. At the national level, an example is Stepping Stones to Caring for Our Children (2013). These are generally determined via a weighting system in licensing or a point system with QRIS.

KEY INDICATORS APPLIED TO OREGON’S CHILD CARE LICENSING SYSTEM

Oregon’s licensing and QRIS systems are very similar to many other states’ licensing and QRIS systems so that the methodologies employed in the past for developing risk assessment and key indicators will be employed in this blueprint. There are some significant challenges because of the psychometric properties of licensing data because of the severe skewness and kurtosis
present in state data systems. These challenges will be addressed later in this blueprint in how to deal skewness and kurtosis.

The risk assessment and key indicators can eventually be tied to the professional development/training/technical assistance system to link resources to specific needs of the programs. It also has the capability of tying them to an early learning benchmarking and child outcomes at some point in the future. This would be accomplished in the full implementation of the Differential Monitoring Logic Model and Algorithm (DMLMA©) as depicted in the Appendix.

TECHNICAL ASPECTS OF THE KEY INDICATOR METHODOLOGY

This section provides the technical and statistical aspects of the key indicator methodology. It will provide the roadmap in taking the Oregon licensing and QRIS data bases through the necessary steps to generating the respective key indicators.

One of the first steps is to sort the data into high and low groups, generally the highest and lowest ratings can be used for this sorting. In very large states this is done on a sampling basis which will be described later in the blueprint. Frequency data will be obtained on those programs in the top level (usually top 20-25%) and the bottom level (usually the bottom 20-25%). The middle levels are not used for the purposes of these analyses. These two groups (top level & the bottom level) are then compared to how each program scored on each item within the specific assessment tool (see Figure 1). An example would be the following: Item 16 from the ECERS – Encouraging Children to Communicate. Sort all the providers by the number in the highest group and the lowest. Then determine how each program scored on item 16, did they get a 5 or higher or a 3 and lower? Fill in the cells within Figure 1 accordingly (see Figure 2).

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest level (top 20-25%)</td>
<td>A</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Lowest level (bottom 20-25%)</td>
<td>C</td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td>Column Total</td>
<td>W</td>
<td>X</td>
<td>Grand Total</td>
</tr>
</tbody>
</table>
Figure 2 depicts that all programs that were in the top 25% (5+ on ECERS, Item 16) were also in the highest rating while the bottom 25% (3 or lower on the ECERS, Item 16) were also in the lowest rating. The data depicted in Figure 2 are taken from studies completed in Pennsylvania in 2002 (Fiene, etal) and 2006 (Barnard, Smith, Fiene & Swanson) in which their quality rating and improvement system (QRIS), Keystone STARS, was validated. The reason for selecting this particular item from the ECERS is that it demonstrates a perfect phi coefficient in discriminating between the highest level and the lowest level. Most, if not all, of the licensing items that will attain the threshold levels to become key indicators will not approach this phi coefficient.

<table>
<thead>
<tr>
<th>Figure 2 – Pa. Study (Fiene, etal, 2002).</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Star level in Pa.</td>
<td>117</td>
<td>0</td>
<td>117</td>
</tr>
<tr>
<td>Lowest Star level in Pa.</td>
<td>0</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Column Total</td>
<td>117</td>
<td>35</td>
<td>152</td>
</tr>
</tbody>
</table>

Once the data are sorted in the above matrix, the following formula (Figure 3) is used to determine if Item 16 is a key indicator or not by calculating its respective Phi coefficient. Please refer back to Figure 1 for the actual placement within the cells and Figure 2 for the data within the cells. The legend (Figure 4) below the formula shows how the cells are defined.

**Figure 3 – Formula for Phi Coefficient**

\[ \phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}} \]

**Figure 4 – Legend for the Cells within the Phi Coefficient**

- **A** = High Group + Programs in Compliance on Specific Compliance Measure.
- **B** = High Group + Programs out of Compliance on Specific Compliance Measure.
- **C** = Low Group + Programs in Compliance on Specific Compliance Measure.
- **D** = Low Group + Programs out of Compliance on Specific Compliance Measure.
- **W** = Total Number of Programs in Compliance on Specific Compliance Measure.
- **X** = Total Number of Programs out of Compliance on Specific Compliance Measure.
- **Y** = Total Number of Programs in High Group.
- **Z** = Total Number of Programs in Low Group.
Once the data are run through the formula in Figure 3, the following chart (Figure 5) can be used to make the final determination of including or not including the item as a key indicator. Based upon the chart in Figure 5, it is best to have a Phi Coefficient approaching +1.00 since we are dealing with normally distributed data. This requirement is relaxed with licensing rules & QRIS selected standards only (+.26 and higher) because the data are more skewed but this should not be the case as much with Oregon’s Quality Rating and Improvement System (QRIS).

Continuing with the chart in Figure 5, if the Phi Coefficient is between +.25 and -.25, this indicates that the indicator is unpredictable in being able to predict overall compliance with the quality rating assessment tool. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance. This can occur with Phi Coefficients above +.25 but it becomes unlikely as we approach +1.00 although there is always the possibility that other standards/rules/regulations could be found out of compliance (this was demonstrated in a study conducted by the author (Fiene, 2013c) with Head Start programs). Another solution is to increase the number of key indicators to be reviewed but this will cut down on the efficiency which is desirable and the purpose of the key indicators.

The last possible outcome with the Phi Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the decision we want to make. The indicator would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously something we do not want to occur.

**Figure 5 – Thresholds for the Phi Coefficient (Fiene & Nixon, 1983, 1985)**

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include</td>
</tr>
<tr>
<td>(+.25) – (-.25)</td>
<td>Unpredictable</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>

The key indicators should then only be used with those programs who have attained the highest rating. It is not intended for those programs that have attained lower ratings. However, even with those programs that have attained the highest rating, every 3-5 years a full, comprehensive review using the full set of rules/standards for licensing and QRIS should occur (see Figure 6 for a graphical depiction). It is intended that a re-validation of the key indicators occur on a periodic basis to make certain that the key indicators have not changed because of differences in compliance history. This is an important and necessary step for the state to engage in to...
ascertain the overall validity and reliability of the assessment system. Also there should not have been any major changes in the program while the key indicators are being administered, such as the director leaving or a large percentage of teachers leaving or enrollment increasing significantly, or a change in the licensing status of the program.

**Figure 6 - Proposed DMLMA System with Key Indicators (KI)**

*Use of Oregon Key Indicators (ORKI) for Licensing and/or QRIS with a Full Review every 4th Year*

![Diagram of proposed DMLMA system with key indicators.](image)

**TECHNICAL ASPECTS OF THE RISK ASSESSMENT METHODOLOGY**

The risk assessment methodology is very different from the key indicator methodology in that compliance history data are not utilized but rather a best practice ranking according to risk is used to determine which rules become core rules which have the greatest likelihood to place children at significant risk of morbidity or mortality. This is done by having a group of experts rank order all the rules on a Likert Scale from low risk to high risk of mortality or morbidity that non-compliance with the rule places children at. This is generally done on a 1-10 scale with 1 = low risk; 5 = medium risk; and 10 = high risk (see Figure 6A). The experts selected include but are not limited to licensing staff, policy makers, researchers, child care providers, advocacy groups, parents, and other significant stakeholders who will be impacted by the weighting of the rules.

**Figure 6A – Example of a Likert Scale for Measuring Risk to Children**

<table>
<thead>
<tr>
<th>Low Risk</th>
<th>Medium Risk</th>
<th>High Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Once the data are collected from all the experts, it is averaged for each rule to determine its relative rank in comparison to all the other rules. A significantly high threshold or cut off point is determined so that no more than 5-10% of the rules become core rules. These core rules can then be used in a differential monitoring approach (to be described more fully in the next section).
and/or with the key indicators to complete abbreviated reviews of child care programs. It is recommended that such a practice of using both core rules and key indicators be used together because than the state has the benefits of both methodologies in measuring risk and being able to statistically predict overall compliance with a very short list of rules.

**TECHNICAL ASPECTS DIFFERENTIAL MONITORING METHODOLOGY**

There are a couple of other key technical aspects that need to be in place for a differential monitoring system to work. The Differential Monitoring Logic Model and Algorithm (DMLMA©)(see the Appendix) is a 4th generational Early Childhood Program Quality Indicator Model4 (ECPQIM4©) in which the major monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With this new model, it is now possible to compare results obtained from licensing systems, quality rating and improvement systems (QRIS), risk assessment systems, key indicator systems, technical assistance, and child development/early learning outcome systems. The various approaches to validation are interposed within this model and the specific expected correlational thresholds that should be observed amongst the key elements of the model are suggested (see Figure 6B).

**Figure 6B – Inter-Correlational Threshold Matrix**

<table>
<thead>
<tr>
<th></th>
<th>PQ</th>
<th>RA</th>
<th>KI</th>
<th>DM</th>
<th>PD</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>PQ</td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>KI</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>DM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
</tbody>
</table>
Key Elements (see the Appendix): CI = state or federal standards, usually rules or regulations that measure health and safety - Caring for Our Children or Head Start Performance Standards will be applicable here. PQ = Quality Rating and Improvement Systems (QRIS) standards at the state level; ERS (ECERS, ITERS, FDCRS), CLASS, or CDPES (Fiene, 2007). RA = risk assessment tools/systems in which only the most critical rules/standards are measured. Stepping Stones is an example of this approach. KI = key indicators in which only predictor rules/standards are measured. The Thirteen Indicators of Quality Child Care is an example of this approach. DM = differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol. PD = technical assistance/training and/or professional development system which provides targeted assistance to the program based upon the DM results. CO = child outcomes which assesses how well the children are developing which is the ultimate goal of the system.

Once the above key elements are in place, it is then possible to look at the relationships amongst them to determine if the system is operating as it was intended. This is done through a validation of the overall system and assessing the inter-correlations (Table 6B) to determine that the DM system is improving the health, safety, program quality and ultimately the overall development of the children it serves.

Oregon should use the following plan to implement the above approach:

STATE AGENCY PLAN (These Steps can be viewed as an overall plan as outlined in Zellman & Fiene (2012):

The first step in utilizing the DMLMA for a state is to take a close look at its Comprehensive Licensing Tool (CI) that it uses to collect violation data on all rules with all facilities in its respective state. If the state does not utilize a tool or checklist or does not review all violation data than it needs to consider these changes because the DMLMA is based upon an Instrument Based Program Monitoring System (IPM) which utilizes tools/checklists to collect data on all rules.

The second step for the state is to compare their state’s rules with the National Health and Safety Performance Standards (Caring for Our Children) to determine the overlap and coverage between the two. This is the first approach to validation which involves Standards review (Zellman & Fiene, 2012).

The third step for the state if it utilizes a Risk Assessment (RA) tool is to assess the relationship between this tool and Stepping Stones to determine the overlap and coverage between the two. This is a continuation of the first approach to validation which involves Standards review (Zellman & Fiene, 2012).
The fourth step for the state is to compare the results from the CI with the RA tools. This step is the second approach to validation which involves Measures (Zellman & Fiene, 2012). The correlation between CI and RA should be at the .50 level or higher (.50+)(see Figure 6B).

In the fifth step, if a state is fortunate enough to have a QRIS – Quality Rating and Improvement System in place and has sufficient program quality (PQ) data available then they will have the ability to compare results from their CI tool with their PQ tool and validate outputs by determining the relationship between compliance with health and safety rules (CI) and program quality (PQ) measures, such as the ERS’s, CLASS, CDPES, etc… This is a very important step because very few empirical demonstrations appear in the research literature regarding this relationship. This step is the third approach to validation which involves Outputs (Zellman & Fiene, 2012). It would be expected that lower correlations (.30+) would be found between CI and PQ because these tools are measuring different aspects of quality such as health & safety versus caregiver-child interactions or overall classroom quality.

The sixth step is for the state to generate a Key Indicator (KI) tool from the CI data base. Please see Fiene & Nixon (1985) and Fiene & Kroh (2000) for a detailed explanation of the methodology for generating a KI tool. This step is also part of the second approach to validation which involves Measures. The correlation between the CI and KI should be very high (.70+) because the KI is a subset of predictor rules taken from the CI data base. If a state did not want to use the KI methodology, a direct comparison could be drawn from The Thirteen Indicators of Quality Child Care (Fiene, 2002).

The seventh step for the state is to use the RA and KI tools together to determine overall compliance of facilities and how often and which rules will be monitored for future visits. This is the basic component of a Differential Monitoring (DM) approach and continues the second approach to validation (Measures). Also, this step should drive decisions within the technical assistance/training/professional development (PD) system in what resources are allocated to a particular facility. It would be expected that moderate correlations (.50+) would be found amongst RA, KI, DM, and PD.

The eighth and final step for the state is to compare the results from the various monitoring tools (CI, PQ, RA, KI) with any child development outcome (CO) data they collect. This is a relatively new area and few, if any, states at this point have this capability on a large scale. However, as Early Learning Networks and Standards are developed, this will become more common place. This step is the forth approach to validation which involves Outcomes (Zellman & Fiene, 2012). The correlations between CI, PQ, RA, KI and CO will be on the lower end (.30+) because there are so many other variables that impact children’s development other than child care facilities.

The last step is to present a logic model which depicts how a differential monitoring system could potentially be actually used in Oregon (see Figure 6C).
Figure 6C – Logic Model for Compliance Decisions

Compliance Decisions:

Core Indicators = Core Rules + Key Indicators – this becomes a screening tool to determine if a program receives an AV or FV visit.

Core Indicators (100%) = The next visit is an Abbreviated Visit. Every 3-4 years a Full Licensing Visit is conducted.

Core Indicators (not 100%) = The next visit is a Full Licensing Visit where all rules are reviewed.

Compliance = 96%+ with all rules which indicates substantial to full compliance with all rules and 100% with Core Indicators. The next visit is an Abbreviated Visit.

Non-compliance = less than 96% with all rules which indicates lower compliance with all rules. The next visit is a Full Visit Study.

SAMPLE

Generally a sample is drawn from the population of early care and education facilities in respective states. Oregon will not be any different because of the size of the overall child care program. A random sample will be selected that represents the state population of child care programs. This will be determined by the number of programs, how the programs are distributed throughout the state, the size of the programs, the type of programs, etc… This will need to be determined once the actual implementation of this blueprint report is started. The author of this report can assist Oregon staff in how best to select the sample of programs.

POTENTIAL CHALLENGES

As mentioned earlier, the measurement issues with licensing data will provide challenges because of their data distributions. In the past when key indicators have been generated with
licensing data which are highly skewed, dichotomization of the data is regularly done\(^3\). Generally dichotomization of data should not be done with normally distributed data\(^4\); however, in this case with QRIS systems, it is appropriate to do so since the data lend themselves to being sorted into discrete categories, such as rating levels. The dichotomization will compare the lowest rating level with the highest rating level in order to generate the key indicators.

**Figure 7 – Data Distribution Comparisons of ERS, QRIS, and Licensing Data**

**TIMELINE**

As soon as all early care and education programs have gone through their assessment phase, it will be possible to do the calculations to determine the Phi Coefficients and generate the key indicators. I am guessing that this should not take any longer than 1 year but could be completed in a much shorter period of time if the assessments on individual programs could be moved up (see Figure 8). The analytical phase should take no longer than a month with an additional month to write up the report. A face to face presentation of the analyses could be done after these two months.

The timeline presented in Figure 8 can be adjusted to the specific needs for the Oregon system. The timeline is based upon previous projects and the average time to generate risk assessment
core rules and key indicators. Another consideration or task is the development of the policies and procedures to be developed and implemented regarding the use of key indicators. This was not specifically listed on the timeline because it is something that is generally developed throughout the project with feedback from all the stakeholders who will be impacted by the use of this new approach to assessment and monitoring.

**Figure 8 - OREGON DMLMA PROJECT TIMELINE**

<table>
<thead>
<tr>
<th>TASK</th>
<th>MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect Data</td>
<td>M1-M3</td>
</tr>
<tr>
<td>Sort Data</td>
<td>M2-3</td>
</tr>
<tr>
<td>Run Analyses</td>
<td>M3-5</td>
</tr>
<tr>
<td>Generate KI/RA</td>
<td>M6</td>
</tr>
<tr>
<td>Train on KI/RA</td>
<td>M6-7</td>
</tr>
<tr>
<td>KI/RA Reliable</td>
<td>M7-9</td>
</tr>
<tr>
<td>Implementation</td>
<td>M10-12</td>
</tr>
</tbody>
</table>

**Legend:**

KI – Key Indicators  
RA – Risk Assessment  
Collect Data – identify participant programs via sampling for KI and the stakeholders for RA.  
Sort Data – KI - the individual programs are sorted into high and low groups representing the top 25% and the bottom 25% of programs as they have scored on the respective rules/standards.  
Run Analyses – KI - each individual item within each of the assessment tools for every program will be compared to the sorting process of the high and low groups. RA – aggregate data into means for each rule, rank order the rules.  
Generate KI/RA – a 2 x 2 matrix is constructed and the key indicators (KI) are generated from this matrix through the use of a phi coefficient. A final report will be delivered to Oregon executive staff for both KI and RA core indicator rules.  
Train on KI/RA – all staff who will be using the KI/RA will be trained on its use.  
KI/RA Reliability – reliability will be established by having two staff go out together and administer the key indicators separately and comparing their results.  
Implementation – once reliability has been established, full implementation will begin.

**COST SAVINGS**

Again based upon previous studies most recently completed in California in 2010 ([http://www.myccl.ca.gov/res/docs/12022010HandoutStakeholderMeeting.pdf](http://www.myccl.ca.gov/res/docs/12022010HandoutStakeholderMeeting.pdf)), time savings of 50% have been attained by using a key indicator or abbreviated tool in completing assessments. It only makes sense that if an assessment can be completed in one hour rather than 2 – 4 hours that a state will see time savings. It is being assumed that equivalent savings should also be the case with Oregon’s licensing/QRIS although this cannot be made certain until the new key indicator or abbreviated tool is actually used for a period of time. Once the new key indicators are used for several months, comparisons could be made to when the full assessments were done.
CONCLUSION AND NEXT STEPS

This blueprint report has given the basic parameters to develop a differential monitoring, risk assessment, and key indicator approach to Oregon’s Licensing/QRIS systems. By following this blueprint Oregon staff should be able to fully implement the approach. Oregon staff would also need to determine if they have the internal capability for the development of the key indicators or if there will be the need to outsource certain aspects of the development. This will be an important consideration as Oregon moves forward with this project. I have provided two options for your consideration in moving forward.

Option 1 – Development of System Internally:

This would require either information systems or research & evaluation staff to analyze the data, generate core key indicator rules, and training of staff. I could provide the necessary consulting services to help the staff work through the methodology. This would probably require at least one face to face meeting with regular monthly conference calls between myself and staff. Discussions of the formatting of data and the types of analyses would be discussed and demonstrated.

Option 2 – Development of System Externally:

In this option I could do all the methodological work demonstrating how I would need the data sent to me, the analytical work in generating core key indicator rules, a report detailing the methodology and results. The only thing that Oregon staff would need to do is get the data to me, all other aspects of what is delineated in the timeline in Figure 8 would be completed by me. This would probably require several face to face trips to explain the process, the results, and do training of staff. Once everything was in place, Oregon staff would have a fully implemented system.

If the above options are of interest I can provide detailed budgets for either one or both.
Notes:

1. The reason for pointing out the need to have a higher Phi Coefficient than what has been reported previously (Fiene & Nixon, 1983, 1985) is the fact that the dichotomization of data should only be used with skewed data and not normally distributed data because it will accentuate differences. However, since the purpose of the dichotomization of data is only for sorting into a high and low group, it would appear to be acceptable for this purpose (MacCallun, etal, 2002. On the practice of dichotomization of quantitative variables, *Psychological Methods, 7, 1*, 19-40.).

2. These results would show an increase in cells B and C in Figure 1 which is undesirable; it should always be the case where $A + D > B + C$ for key indicators to maintain their predictive validity.

3. The distinction between making decisions with skewed (Licensing) as versus normally distributed (ERS) data is an important one because there is a greater likelihood with skewed data of introducing less than optimal programs into the high group when sorting programmatic data into high and low groups. This then makes it more difficult to identify the best programs. However, because of the distribution with skewed data the same cannot be said with the low group in which case it is relatively easy to identify the problem programs. This is not as much of a concern when the data are more normally distributed in which it is relatively easy to identify both the optimal and problem programs. This is an excellent example of the need of weighting of standards in order to increase the normal distribution of the data.
REFERENCES AND ADDITIONAL RELATED READINGS REGARDING DMLMA:


☐ Fiene (2013d). Kansas Child Care Key Indicators. Middletown: Pennsylvania, Research Institute for Key Indicators.


☐ Fiene (2002b). Improving child care quality through an infant caregiver mentoring project, Child and Youth Care Forum, 31(2), 75-83.


For additional information regarding this report, please contact:
Richard Fiene, Ph.D., Director/President
Research Institute for Key Indicators (RIKI)
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
RIKI.Institute@gmail.com
717-944-5868 Phone and Fax
http://RIKIInstitute.wikispaces.com
http://pennstate.academia.edu/RickFiene
**Appendix**

DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA®) (Fiene, 2012): A 4\textsuperscript{th} Generation ECPQIM – Early Childhood Program Quality Indicator Model

\[ \text{CI x PQ} \Rightarrow \text{RA} + \text{KI} \Rightarrow \text{DM} + \text{PD} \Rightarrow \text{CO} \]

**Definitions of Key Elements:**

- **PC** = Program Compliance/Licensing (Health and Safety) (*Caring for Our Children*)
- **PQ** = QRIS/Accreditation/Caregiver/Child Interactions/Classroom Environment Quality (*ERS/CLASS/PAS/BAS*)
- **RA** = Risk Assessment, (High Risk Rules) (*Stepping Stones*)
- **KI** = Key Indicators (Predictor Rules) (*13 Key Indicators of Quality Child Care*)
- **DM** = Differential Monitoring (How often to visit and what to review)
- **PD** = Professional Development/Technical Assistance/Training (Not pictured but part of Model)
- **CO** = Child Outcomes (Not pictured but part of Model)

---

[Diagram of the logic model flowchart]
Qualistar Rating Key Indicator Study

Richard Fiene, Ph.D.

June 17, 2014

ABSTRACT

This report provides an analysis of Colorado’s quality rating system, the Qualistar Rating, for generating key indicators. Key indicators have been used a great deal in the licensing literature but this is a first time analysis in utilizing this methodology in a QRS (Quality Rating System) or a QRIS (Quality Rating and Improvement System). The key indicator methodology is described in detail applying it to QRS/QRIS. The results clearly indicate that the strongest key indicators are within the Family Partnerships component of the Qualistar Rating; however there are some major limitations to utilizing this methodology with QRS/QRIS.

INTRODUCTION

The Qualistar Rating, administered by Qualistar Colorado, is one of the longest continuously running QRS in the United States. Presently over 50% of states have QRS/QRIS and the research on these program quality rating & improvement systems has increased over the years. One area of research that has been gaining momentum most recently is ascertaining the most effective and efficient delivery system for a QRS/QRIS as the number of early care and education programs participating in QRS/QRIS continues to increase. This report provides an overview to the topic and introduces an option that has been used in the human services/child care licensing field in identifying key indicators of overall compliance with standards. The purpose of the key indicator methodology is to focus monitoring visits on those standards that have the ability to predict overall compliance with the full set of QRS/QRIS standards. The key indicator methodology is part of a program monitoring approach called Differential Program Monitoring which was developed to help streamline the program monitoring of early care and education programs (please see the Appendix for two graphics which help to depict this relationship (Figures 8/9)). It was first applied in child care licensing (Fiene & Nixon, 1985) but has been used in many other service types, such as: Head Start Performance Standards (Fiene,
2013a), National Accreditation (Fiene, 1996), and child and adult residential programs (Kroh & Melusky, 2010). The methodologies are based upon statistical protocols that have been developed in the tests and measurements literature in which an abbreviated set of items is used to statistically predict as if the full test was applied. This methodology has been used in regulatory analysis and is now being proposed for use in Quality Rating and Improvement Systems (Fiene, 2013b). This study and report is the first demonstration of its use with QRS.

TECHNICAL ASPECTS OF THE KEY INDICATOR METHODOLOGY

This section provides the technical and statistical aspects of the key indicator methodology. It will provide the specific methodology for generating the key indicators for the Qualistar Rating.

One of the first steps is to sort the data into high and low groups, generally the highest and lowest ratings can be used for this sorting. In very large states such as Colorado this is done on a sampling basis. Frequency data will be obtained on those programs in the top level (usually top 20-25%) and the bottom level (usually the bottom 20-25%). The middle levels are not used for the purposes of these analyses. These two groups (top level & the bottom level) are then compared to how each program scored on each item within the specific assessment tool (see Figure 1). An example from the Qualistar Rating database is provided in Figure 2 (see Figure 2).

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highest level</strong>&lt;br&gt;(top 20-25%)</td>
<td>A</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Lowest level</strong>&lt;br&gt;(bottom 20-25%)</td>
<td>C</td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td><strong>Column Total</strong></td>
<td>W</td>
<td>X</td>
<td>Grand Total</td>
</tr>
</tbody>
</table>

Because of the differences in the data distribution for the Qualistar Rating, the above cutoff points had to be more stringent with the respective cutoff points for the high and low groups because the majority of the programs were at the Star 2 and 3 levels. In comparing these data to past licensing distributions (see Fiene, 2013d), it would be expected that the majority of programs would be at a Star 1 level, but that was not the case with this sample. Rather than using a 20-25% cut off point, it was changed to 10% to accommodate this difference. Figure 2 depicts that all programs that were in the top 10% were in the highest rating while the bottom 10% were in the lowest rating. The data depicted in Figure 2 are taken from the Family
Engagement Standard 5 – The program provides opportunities for staff and families to get to know one another. The reason for selecting this particular standard is that it demonstrates a perfect Phi Coefficient in discriminating between the highest level and the lowest level.1

<table>
<thead>
<tr>
<th>Figure 2: Criterion 5 Family Partnerships</th>
<th>Providers In Compliance or Top 10%1</th>
<th>Programs Out Of Compliance or Bottom 10%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Star level</td>
<td>11</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Lowest Star level</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Column Total</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
</tbody>
</table>

Once the data are sorted in the above matrix, the following formula (Figure 3) is used to determine if the standard is a key indicator or not by calculating its respective Phi Coefficient. Please refer back to Figure 1 for the actual placement within the cells and Figure 2 for the data within the cells. The legend (Figure 4) below the formula shows how the cells are defined.

Figure 3 – Formula for Phi Coefficient

\[ \phi = \frac{(A)(D)-(B)(C)}{\sqrt{(W)(X)(Y)(Z)}} \]

Figure 4 – Legend for the Cells within the Phi Coefficient

\[ A = \text{High Group + Programs in Compliance on Specific Compliance Measure.} \]
\[ B = \text{High Group + Programs out of Compliance on Specific Compliance Measure.} \]
\[ C = \text{Low Group + Programs in Compliance on Specific Compliance Measure.} \]
\[ D = \text{Low Group + Programs out of Compliance on Specific Compliance Measure.} \]
\[ W = \text{Total Number of Programs in Compliance on Specific Compliance Measure.} \]
\[ X = \text{Total Number of Programs out of Compliance on Specific Compliance Measure.} \]
\[ Y = \text{Total Number of Programs in High Group.} \]
\[ Z = \text{Total Number of Programs in Low Group.} \]

Once the data are run through the formula in Figure 3, the following chart (Figure 5) can be used to make the final determination of including or not including the item as a key indicator. Based
upon the chart in Figure 5, it is best to have a Phi Coefficient approaching +1.00 since the data are more normally distributed than is the case with licensing data.

Continuing with the chart in Figure 5, a Phi Coefficient between +.75 and -.25 indicates that the indicator is unpredictable in being able to predict overall compliance with the quality rating assessment tool. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance. This can occur with Phi Coefficients above +.75 but it becomes unlikely as they approach +1.00, although there is always the possibility that other standards/rules/regulations could be found to be out of compliance (this was demonstrated in a study conducted by the author (Fiene, 2013c). Another solution is to increase the number of key indicators to be reviewed but this will cut down on the efficiency which is desirable and the purpose of the key indicators.

The last possible outcome with the Phi Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the desired. The indicator would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously undesirable.

**Figure 5 – Thresholds for the Phi Coefficient (Fiene & Nixon, 1983, 1985)(Fiene, 2014)**

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.76)</td>
<td>Good Predictor</td>
<td>Include</td>
</tr>
<tr>
<td>(+.75) – (-.25)</td>
<td>Unpredictable</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>

The key indicators should then only be used with those programs that have attained the highest rating. It is not intended for those programs that have attained lower ratings. However, even with those programs that have attained the highest rating, periodically a full, comprehensive review using the full set of standards for Qualistar Colorado should occur (see Figure 6 for a graphical depiction). It is intended that a re-validation of the key indicators occur on a periodic basis to make certain that the key indicators have not changed because of differences in compliance with standards history. This is an important and necessary step for the program to engage in to ascertain the overall validity and reliability of the assessment system. Also there should not have been any major changes in the program while the key indicators are being administered, such as the director leaving or a large percentage of teachers leaving or enrollment increasing significantly, or a change in the licensing or accreditation status of the program.
**RESULTS**

The results reported in this section are based upon a sample selected from the overall Qualistar Rating database from its most recent monitoring reviews (N = 117). This was a representative sample of the program’s QRS.

There are five components of the Qualistar Rating: Learning Environment, Family Partnerships, Training and Education, Adult to Child Ratios and Group Size, and Accreditation. See Figures 10-14 in the Appendix for the graphical depictions of the data distributions for the five major criteria. The data distributions are provided because a pre-requisite for calculating the key indicator Phi Coefficients is the dichotomization of data with a skewed data distribution. Figures 10-14 display how much the data are skewed.

The Qualistar Rating is a zero-to-4 star system, with 4 stars indicating the highest level of quality\(^4\). Eleven programs were rated at the Star 1 level, 19 programs were rated at the Star 2 level, 77 programs were rated at the Star 3 level, and 10 programs were rated at the Star 4 level for a total of 117 programs included in these analyses. There were no programs in the sample that earned less than one star.

Based upon the key indicator methodology described in the previous section, the only Qualistar Rating standards that reached key indicator designation\(^5\) were the following: *Family Partnership Standard/Criterion 5 = The program provides opportunities for staff and families to get to know one another; Family Partnership Standard/Criterion 7 = Families receive information on their child’s progress on a regular basis, using a formal mechanism such as a report or parent conference and Family Partnership Standard/Criterion 8 = Families are included in planning and decision making for the program.*
Figure 7 – Key Indicators with Phi Coefficients

<table>
<thead>
<tr>
<th>Family Partnership Standard/Criterion</th>
<th>Phi</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.00</td>
<td>.001</td>
</tr>
<tr>
<td>7</td>
<td>0.86</td>
<td>.001</td>
</tr>
<tr>
<td>8</td>
<td>0.83</td>
<td>.001</td>
</tr>
</tbody>
</table>

There were many other significant correlations (Family Partnerships and Adult-to-Child Ratios and Group Sizes) obtained but none reached the cutoff threshold of .76+ for the Phi calculations. These other correlations are reported in the Appendix after the descriptive graphical displays in Figures 15, 15a, 15b. The Phi Coefficients for the other Criteria (Learning Environment, Training and Education, and Program Accreditation) were not calculated because the data distributions were not skewed as was the case with Family Partnerships and Adult-to-Child Ratios and Group Sizes (see Figures 10-14).

LIMITATIONS

There are two major limitations to this study, 1) the first deals with the statistics being used to generate the key indicators; 2) the second deals with the key indicator methodology.

The first limitation has to do with dichotomization of data which should only be used with very skewed data. Data skewness always occurs with licensing data because of the nature of the data, health and safety protections (the majority of programs are always in compliance with the respective rules). However, this appears to not always be the case with QRS/QRIS data which deals with more program quality aspects of facilities and shows greater variation in the data. If this is the case then dichotomization of data is not appropriate and should not be utilized in order to generate key indicators.

The second limitation of this study is if the key indicator methodology and differential monitoring approaches are appropriate for QRS/QRIS. In Figure 6 above and in the conclusion to this report below, there is a scenario where it can be used but Qualistar Colorado and each state must determine if this is an appropriate approach for their respective program. For example, key indicators will not work in a block model and with a point-system model may generate very limited time savings if the data distribution is normally distributed and there are very few programs at the highest star level. In licensing data base distributions there is always a large number of programs to select from in the highest compliance levels (usually a minimum of 25%).
CONCLUSION/FUTURE RESEARCH/DISCUSSION/RECOMMENDATIONS

This study is the first of its kind in generating key indicators for a QRS based upon the analyses performed with the Qualistar Rating data base. It potentially demonstrates that the use of the key indicator methodology with QRS/QRIS could be feasible and warranted in order to focus limited program monitoring resources in a most efficient and effective manner keeping the above stated limitations in mind as stated in the previous Limitations Section. In the future, Qualistar Colorado may want to pilot an approach utilizing a small group of programs and could focus resources on the Family Partnership/Engagement standards on an ongoing basis between comprehensive reviews as depicted in Figure 6 above for Star 4 programs. The time saved here could then be redistributed to spending more time with the Star 1 programs.

It will be timely to see other states and programs who are interested in generating key indicators if they have Family Partnership/Engagement standards as part of their respective QRS/QRIS to determine if these standards reach the same threshold for key indicator designation as has occurred in this study. It will also be interesting to see if any other state’s criteria/standards data distributions are similar to what has been found in the Qualistar Rating or not.

However, as highlighted in the Limitations Section, states and programs need to consider if the key indicator methodology and the resultant differential monitoring model is really warranted and appropriate for their respective QRS/QRIS’s. As has been the case with Colorado’s Qualistar Rating, only two of the five major criteria: Family Partnerships and Adult-Child Ratio/Group Size were determined to be good candidates for the key indicator Methodology in which the data were skewed enough to warrant dichotomization. The other three major criteria: Learning Environment, Training and Education, and Program Accreditation were determined not to be sufficiently skewed to warrant dichotomization. This sets up a decision making system in which only 40% of the criteria are being used and severely limits the overall predictability of the key indicators selected. Could the other criteria be used to generate key indicators? Of course, but dichotomization of data should not be done when data are not highly skewed (MacCallun, etal, 2002). Yes, we were successful in generating Key Indicators for the Qualistar Rating but within a limited scenario in how they should be used. The results are not equivalent to what has been found and utilized in the licensing literature where the licensing data are always highly skewed. If a state or program find that all the standards are skewed in a similar way to licensing data then dichotomization of data and the generation of key indicators is warranted.

A recommendation to Colorado’s Qualistar and other programs and states where they find the data from their standards more normally distributed that they not use a key indicator approach. The key indicator approach remains a reliable and valid methodology for licensing but only in very special and limited cases will it be an appropriate monitoring approach for more program quality focused systems, such as QRS/QRIS and accreditation. For those QRS/QRIS systems where the standards are more normally distributed, the recommendation would be to continue to use the full set of QRS/QRIS standards and not use an abbreviated set of standards.
NOTES:

1. For analytical purposes, the top 10% of programs received an average score of 8 points or higher on a 10 point scale and the bottom 10% of programs received an average score of 2 points or less on a 10 point scale.

2. The reason for pointing out the need to have a higher Phi Coefficient than what has been reported previously (Fiene & Nixon, 1983, 1985) is the fact that the dichotomization of data should only be used with skewed data and not normally distributed data because it will accentuate differences. However, since the purpose of the dichotomization of data is only for sorting into a high and low group, it would appear to be acceptable for this purpose (MacCallun, etal, 2002. On the practice of dichotomization of quantitative variables, Psychological Methods, 7, 1, 19-40.).

3. These results would show an increase in cells B and C in Figure 1 which is undesirable; it should always be the case where A + D > B + C for key indicators to maintain their predictive validity.

4. The following point values equate to the various Star levels in the Qualistar Rating System (for detailed information regarding the QRS system please see the following document: Qualistar Colorado – Qualistar Rating Criteria Chart, November 2012):
   - Provisional = 0 – 9 points or Learning Environment score of 0
   - Star 1 = 10 - 17 points
   - Star 2 = 18 - 25 points
   - Star 3 = 26 - 33 points
   - Star 4 = 34 - 42 points

   Qualistar Rating Criteria Chart:
   - Learning Environment = points are awarded based on average classroom scores on the ERS Scales. (Score of component: 1 – 10)
   - Family Partnerships = points are awarded based on how well programs communicate with collaborate with, and involve families. Score of component: 1 – 10)
   - Training and Education = points are awarded to teachers & center administrators based on their professional development level and amount of experience, with criteria separated by position. Score of component: 1 – 10
   - Adult-to-Child Ratios & Group Size = points are awarded based on the average adult-to-child ratio and group size in each classroom. Score of component: 1 – 10
   - Program Accreditation = points are awarded for receiving and maintaining national program accreditation through an approved organization. Score of component: 0 or 2 points

   The reader needs to keep in mind that Qualistar Colorado is not a state agency but rather a private non-profit agency.

5. The three Family Partnership Standards were met at the Star 4 level always or most of the time (see Figure 2).

6. The respective skewness figures are the following: Family Partnership = -1.425; Adult-Child Ratio/Group Size = -1.506; Learning Environment = -0.946; Training and Education = 0.028; Program Accreditation = 7.548. See Figure 16 for basic descriptive statistics for these Criteria.

For additional information regarding this Report, please contact:
Richard Fiene, Ph.D., Director/President, Research Institute for Key Indicators (RIKI), 41 Grandview Drive, Middletown, PA. 17057; DrFiene@gmail.com; 717-944-5868 Phone and Fax; http://RIKInstitute.wikispaces.com
REFERENCES AND ADDITIONAL RELATED READINGS REGARDING DIFFERENTIAL MONITORING, RISK ASSESSMENT, AND KEY INDICATOR METHODOLOGIES:


Appendix – Figure 8

DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA©) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model

CI x PQ => RA + KI => DM + PD => CO

Definitions of Key Elements:

PC = Program Compliance/Licensing (Health and Safety) (Caring for Our Children)
PQ = QRIS/Accreditation/Caregiver/Child Interactions/Classroom Environment Quality (ERS/CLASS/PAS/BAS)
RA = Risk Assessment, (High Risk Rules) (Stepping Stones)
KI = Key Indicators (Predictor Rules) (13 Key Indicators of Quality Child Care)
DM = Differential Monitoring (How often to visit and what to review)
PD = Professional Development/Technical Assistance/Training (Not pictured but part of Model)
CO = Child Outcomes (Not pictured but part of Model)
Appendix – Figure 9 - Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

All Licensing Rules – Full Compliance Reviews

Differential Monitoring

How Often to Visit? What is Reviewed?

Frequency

More Often Less Often

Abbreviated Tool

Risk Assessment Weights Key Indicators Predictors
Figures 10-14 depict the data distributions for overall Star points as well as for the major criteria/standards (Training & Education, Learning Environment, Adult-to-Child Ratios & Group Size, and Family Partnerships). Figures 13-14 clearly demonstrate how these respective criteria/standards are extremely skewed data distributions while Figures 10-12 show a more normally distributed data pattern. This is important for which standards can be dichotomized and phi coefficients generated. Dichotomization of data should only be used with skewed data which is the case in figures 13-14. It is not appropriate with the data distributions in figures 10-12. Also see Figure 16 for additional descriptive statistics for the specific criteria.
Figure 11
Figure 12

LEARNING ENVIRONMENT POINTS
Figure 14

FAMILY PARTNERSHIP STAR POINTS

0 2 4 6 8 10

0 10 20 30 40 50 60 70
Figure 15

Selected Relationships amongst the Standards/Criteria and Star Level

<table>
<thead>
<tr>
<th>Standards/Criteria</th>
<th>Correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Partnerships x Star Level</td>
<td>.80****</td>
</tr>
<tr>
<td>Learning Environment x Star Level</td>
<td>.68***</td>
</tr>
<tr>
<td>Training/Education x Star Level</td>
<td>.54**</td>
</tr>
<tr>
<td>Adult-Child Ratio/Group Size x Star Level</td>
<td>.46*</td>
</tr>
<tr>
<td>Program Accreditation x Star Level</td>
<td>.11</td>
</tr>
</tbody>
</table>

* p < .05
** p < .01
*** p < .001
**** p < .0001

Figure 15a

<table>
<thead>
<tr>
<th>Family Partnership Criteria</th>
<th>Phi</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 1</td>
<td>.23</td>
<td>ns</td>
</tr>
<tr>
<td>Criterion 2</td>
<td>.53</td>
<td>.02</td>
</tr>
<tr>
<td>Criterion 3</td>
<td>.46</td>
<td>.04</td>
</tr>
<tr>
<td>Criterion 4</td>
<td>.46</td>
<td>.04</td>
</tr>
<tr>
<td>Criterion 5</td>
<td>1.00</td>
<td>.001</td>
</tr>
<tr>
<td>Criterion 6</td>
<td>.46</td>
<td>.04</td>
</tr>
<tr>
<td>Criterion 7</td>
<td>.86</td>
<td>.001</td>
</tr>
<tr>
<td>Criterion 8</td>
<td>.83</td>
<td>.001</td>
</tr>
<tr>
<td>Criterion 9</td>
<td>.72</td>
<td>.001</td>
</tr>
<tr>
<td>Criterion 10</td>
<td>.60</td>
<td>.006</td>
</tr>
<tr>
<td>Criterion 11</td>
<td>.46</td>
<td>.04</td>
</tr>
<tr>
<td>Criterion 12</td>
<td>.53</td>
<td>.02</td>
</tr>
<tr>
<td>Criterion 13</td>
<td>.21</td>
<td>ns</td>
</tr>
<tr>
<td>Criterion 14</td>
<td>.46</td>
<td>.04</td>
</tr>
<tr>
<td>Criterion 15</td>
<td>.39</td>
<td>ns</td>
</tr>
<tr>
<td>Criterion 16</td>
<td>.75</td>
<td>.001</td>
</tr>
<tr>
<td>Criterion 17</td>
<td>.60</td>
<td>.006</td>
</tr>
</tbody>
</table>

**Legend:**
Criteria 1 – 7 involve the program providing information to families.
Criteria 8 – 15 involve families in planning, communicating and decision making for the program.
Criteria 16 – 17 involve a written plan and evaluating the program’s family partnerships.
Figure 15b

<table>
<thead>
<tr>
<th>Adult-Child Ratio/Group Size</th>
<th>Phi</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult-Child Ratios</td>
<td>.58</td>
<td>.0001</td>
</tr>
<tr>
<td>Group Size</td>
<td>.33</td>
<td>.02</td>
</tr>
</tbody>
</table>

Family Partnerships and Adult-Child Ratio/Group Size standards/criteria phi coefficients were generated because of the skewed data distributions. Phi coefficients were not generated for Learning Environment, Training and Education or Program Accreditation because the data were not sufficiently skewed or showed no variability at all in their respective distributions.

Figure 16

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Mean</th>
<th>Median</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Partnerships</td>
<td>7.7</td>
<td>10</td>
<td>-1.425</td>
</tr>
<tr>
<td>Adult-to-Child Ratios &amp; Group Size</td>
<td>9.1</td>
<td>10</td>
<td>-1.506</td>
</tr>
<tr>
<td>Learning Environment</td>
<td>5.8</td>
<td>6</td>
<td>-0.946</td>
</tr>
<tr>
<td>Training and Education</td>
<td>4.7</td>
<td>5</td>
<td>0.028</td>
</tr>
<tr>
<td>Program Accreditation</td>
<td>0.0</td>
<td>0</td>
<td>7.548</td>
</tr>
<tr>
<td>Total Star Level</td>
<td>2.7</td>
<td>3</td>
<td>-1.213</td>
</tr>
</tbody>
</table>
OREGON’S STEPPING STONES\textsuperscript{1} RISK FACTORS ANALYSIS

The purpose of this analysis is to provide Oregon OCC with a basic risk factor analysis comparing its child care center rules to Stepping Stones (SS) standards. This analysis will delineate, based upon Stepping Stones’ major content areas (chapters from Caring for our Children (CFOC)), where there may be gaps in their child care center rules.

This analysis is a summary look at the comparison between Stepping Stones and Oregon’s Rules; it is now intended to be an in-depth crosswalk between the two sets of standards and rules. In order to do that type of analysis, Fiene’s Stepping Stones to Validate State Rules Template (2013) is the suggested source to use.

Table 1 provides the comparisons between Stepping Stones and the Oregon Child Care Center Rules in which a search of the rules was done to determine if the specific SS standard was present or not. Every time the search contained a match, it was recorded as a “1”. When there was no match, it was recorded as a “0”.

Table 1 – Comparison of Stepping Stones (SS) Standards and Oregon Child Care Center Rules

<table>
<thead>
<tr>
<th>SS</th>
<th>RULES</th>
<th>PERCENT</th>
<th>CONTENT AREA/RISK FACTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>11</td>
<td>79</td>
<td>STAFFING</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>56</td>
<td>PROGRAM ACTIVITIES FOR HEALTHY DEVELOPMENT</td>
</tr>
<tr>
<td>25</td>
<td>16</td>
<td>64</td>
<td>HEALTH PROMOTION/PROTECTION</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>77</td>
<td>NUTRITION AND FOOD SERVICE</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>60</td>
<td>FACILITIES, SUPPLIES, EQUIPMENT, ENVIRON HEALTH</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>33</td>
<td>PLAY AREAS/PLAYGROUNDS AND TRANSPORTATION</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>INFECTIOUS DISEASES</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>70</td>
<td>POLICIES</td>
</tr>
<tr>
<td>122</td>
<td>69</td>
<td>56.125</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

Legend for Table 1:
Nominal scaling to determine if the Oregon CCC Rules have any reference to the specific SS3 Standard.
It is scored 1/0 where 1 = Present and 0 = Absent. Percent is the total number of “1”. Higher the percent the better.

SS = STEPPING STONES STANDARDS
RULES = OREGON CHILD CARE CENTER RULES
PERCENT = RULES/SS
CONTENT = RISK FACTOR/SS/CFOC CHAPTER
This comparison was completed on the major chapter headings in *Stepping Stones* and *Caring for our Children* as delineated in the Content/Risk Factor Column in Table 1. The following table (Table 2) provides the detail of the contents of each content area/risk factor.

**Table 2 – Major Content/Risk Factor Areas (1-8) and Specific Content for Each Area**

<table>
<thead>
<tr>
<th>1. STAFFING</th>
<th>A. CHILD:STAFF RATIO AND GROUP SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B. RECRUITMENT AND BACKGROUND</td>
</tr>
<tr>
<td></td>
<td>SCREENING</td>
</tr>
<tr>
<td></td>
<td>C. DIRECTOR’S QUALIFICATIONS</td>
</tr>
<tr>
<td></td>
<td>D. TEACHER’S QUALIFICATIONS</td>
</tr>
<tr>
<td></td>
<td>E. PRE-SERVICE TRAINING</td>
</tr>
<tr>
<td></td>
<td>F. ORIENTATION TRAINING</td>
</tr>
<tr>
<td></td>
<td>G. FIRST AID AND CPR TRAINING</td>
</tr>
<tr>
<td></td>
<td>H. STAFF HEALTH</td>
</tr>
<tr>
<td>2. PROGRAM ACTIVITIES FOR HEALTH DEVELOPMENT</td>
<td>A. PROGRAM ACTIVITIES FOR INFANTS,</td>
</tr>
<tr>
<td></td>
<td>TODDLERS, PRESCHOOLERS, AND SCHOOL</td>
</tr>
<tr>
<td></td>
<td>AGE CHILDREN</td>
</tr>
<tr>
<td></td>
<td>B. SUPERVISION AND DISCIPLINE</td>
</tr>
<tr>
<td></td>
<td>C. HEALTH INFORMATION SHARING</td>
</tr>
<tr>
<td></td>
<td>D. HEALTH EDUCATION FOR CHILDREN</td>
</tr>
<tr>
<td></td>
<td>E. HEALTH EDUCATION FOR STAFF</td>
</tr>
<tr>
<td></td>
<td>F. HEALTH EDUCATION FOR PARENTS</td>
</tr>
<tr>
<td>3. HEALTH PROMOTION AND PROTECTION</td>
<td>A. DAILY HEALTH CHECK</td>
</tr>
<tr>
<td></td>
<td>B. ROUTINE HEALTH SUPERVISION</td>
</tr>
<tr>
<td></td>
<td>C. PHYSICAL ACTIVITY AND LIMITING</td>
</tr>
<tr>
<td></td>
<td>SCREEN TIME</td>
</tr>
<tr>
<td></td>
<td>D. SAFE SLEEP</td>
</tr>
<tr>
<td></td>
<td>E. ORAL HEALTH</td>
</tr>
<tr>
<td></td>
<td>F. DIAPERING AND CHANGING SOILED</td>
</tr>
<tr>
<td></td>
<td>CLOTHING</td>
</tr>
<tr>
<td></td>
<td>G. HAND HYGIENE</td>
</tr>
<tr>
<td></td>
<td>H. EXPOSURE TO BODY FLUIDS</td>
</tr>
<tr>
<td></td>
<td>I. EMERGENCY PROCEDURES</td>
</tr>
<tr>
<td></td>
<td>J. CHILD ABUSE AND NEGLECT</td>
</tr>
<tr>
<td></td>
<td>K. INCLUSION/EXCLUSION DUE TO ILLNESS</td>
</tr>
<tr>
<td></td>
<td>L. CARING FOR CHILDREN WHO ARE ILL</td>
</tr>
<tr>
<td></td>
<td>M. MEDICATIONS</td>
</tr>
<tr>
<td>4. NUTRITION AND FOOD SERVICE</td>
<td>A. MEAL SERVICE, SEATING, SUPERVISION</td>
</tr>
<tr>
<td></td>
<td>B. FOOD BROUGHT FROM HOME</td>
</tr>
<tr>
<td></td>
<td>C. KITCHEN AND EQUIPMENT</td>
</tr>
<tr>
<td></td>
<td>D. FOOD SAFETY</td>
</tr>
</tbody>
</table>
| 5. FACILITIES, SUPPLIES, EQUIPMENT, AND ENVIRONMENTAL HEALTH | E. MEALS FROM OUTSIDE VENDORS OR CENTRAL KITCHEN  
F. NUTRITION LEARNING EXPERIENCES FOR CHILDREN  
G. NUTRITION EDUCATION FOR PARENTS |
|---|---|
| A. GENERAL LOCATION, LAYOUT, AND CONSTRUCTION OF THE FACILITY  
B. SPACE PER CHILD  
C. EXITS  
D. STEPS AND STAIRS  
E. EXTERIOR AREAS  
F. VENTILATION, HEATING, COOLING, AND HOT WATER  
G. LIGHTING  
H. NOISE  
I. ELECTRICAL FIXTURES AND OUTLETS  
J. FIRE WARNING SYSTEMS  
K. WATER SUPPLY AND PLUMBING  
L. SEWAGE AND GARBAGE  
M. INTEGRATED PEST MANAGEMENT  
N. PREVENTION AND MANAGEMENT OF TOXIC SUBSTANCES  
O. TOILET AND HANDWASHING AREAS  
P. DIAPER CHANGING AREAS  
Q. SLEEP AND REST AREAS |
| 6. PLAY AREAS/PLAYGROUNDS AND TRANSPORTATION | A. PLAYGROUND SIZE AND LOCATION  
B. USE ZONES AND CLEARANCE REQUIREMENTS  
C. PLAY AREA AND PLAYGROUND SURFACING  
D. INSPECTION OF PLAY AREAS AND EQUIPMENT  
E. ACCESS TO AND SAFETY AROUND BODIES OF WATER  
F. POOL EQUIPMENT AND MAINTENANCE  
G. WATER QUALITY OF POOLS  
H. TRANSPORTATION SAFETY |
| 7. INFECTIOUS DISEASES | A. HOW INFECTIONS SPREAD  
B. IMMUNIZATIONS  
C. RESPIRATORY TRACT INFECTIONS  
D. ENTERIC (DIARRHEAL) INFECTIONS AND HEPATITIS A VIRUS (HAV)  
E. SKIN AND MUCOUS MEMBRANE INFECTIONS |
Table 2 provides you with the specific content as it relates to the risk factors. Figures 1 and 2 as well as Table 3 will provide the comparison between SS standards and Oregon’s child care center rules by these content areas/risk factors.

Figure 1 does this comparison by listing for each content area/risk factor the frequency count where there is a match between rules and standards.

**Figure 1 – Comparing Stepping Stones (SS) Standards and Oregon’s Child Care Center Rules**

<table>
<thead>
<tr>
<th>Legend for Figure 1:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = STAFFING</td>
</tr>
<tr>
<td>2 = PROGRAM ACTIVITIES FOR HEALTHY DEVELOPMENT</td>
</tr>
<tr>
<td>3 = HEALTH PROMOTION/PROTECTION</td>
</tr>
<tr>
<td>4 = NUTRITION AND FOOD SERVICE</td>
</tr>
</tbody>
</table>
Figure 2 takes the data from Table 1 and Figure 1 and expresses the content areas/risk factors in the form of percents in which the percents represent the number of times the Oregon child care center rules and the Stepping Stones standards match.

Legend for Figure 1:
1 = STAFFING
2 = PROGRAM ACTIVITIES FOR HEALTHY DEVELOPMENT
3 = HEALTH PROMOTION/PROTECTION
4 = NUTRITION AND FOOD SERVICE
5 = FACILITIES, SUPPLIES, EQUIPMENT, ENVIRON HEALTH
6 = PLAY AREAS/PLAYGROUNDS AND TRANSPORTATION
7 = INFECTIOUS DISEASES
8 = POLICIES

It is evident from Table 1 and Figures 1 and 2 that the two areas where the greatest gap between the Stepping Stones standards and Oregon’s child care center rules is in the Infectious Diseases and Play Areas/Playgrounds and Transportation content areas/risk factors with a match rate of 10% and 33% respectively. The highest match rates are with the Staffing (79%) and Nutrition & Food Service (77%).
Based upon the above results there are some recommendations to be made where Oregon Office of Child Care staff may want to focus their attention for future rule formulation in the infectious diseases and the play area/playgrounds & transportation content areas.

**Notes:**
1. The reason for using *Stepping Stones* rather than *Caring for our Children* is that *Stepping Stones* are the selected standards from *CFOC* that place children at greatest risk of mortality and morbidity if the standards are not complied with.

---

*For additional information regarding this report, please contact:*

Richard Fiene, Ph.D., President, Research Institute for Key Indicators LLC (RIKI)
41 Grandview Avenue, Middletown, Pennsylvania 17057
Phone/Fax: 717-944-5868; Mobile: 717-598-8908
Website: rikinstitute.wikispaces.com
Email: RIKI.Institute@gmail.com
NARA Illinois Key Indicator Report for Centers, Group Homes, and Family Homes

Richard Fiene, Ph.D.

May 30, 2014

ABSTRACT

This report will provide an analysis of Illinois Rules for child care centers, group homes, and family homes for generating key indicators. There is a brief introduction regarding differential monitoring and key indicators followed by the generated key indicators.

INTRODUCTION

The key indicator methodology is part of a program monitoring approach called Differential Program Monitoring which was developed to help streamline the program monitoring of early care and education programs (please see the appendix for two graphics which help to depict this relationship). It was first applied in child care licensing but has been used in many other service types, such as: Head Start Performance Standards, National Accreditation, and child and adult residential programs. The methodologies are based upon statistical protocols that have been developed in the tests and measurements literature in which an abbreviated set of items is used to statistically predict as if the full test was applied. This methodology has been used in regulatory analysis and is now being proposed for use in Quality Rating and Improvement Systems (QRIS).

TECHNICAL ASPECTS OF THE KEY INDICATOR METHODOLOGY

This section provides the technical and statistical aspects of the key indicator methodology. One of the first steps is to sort the data into high and low groups, generally the highest and lowest ratings can be used for this sorting. In very large states this is done on a sampling basis which
will be described later in the blueprint. Frequency data will be obtained on those programs in the top level (usually top 20-25%) and the bottom level (usually the bottom 20-25%). The middle levels are not used for the purposes of these analyses. These two groups (top level & the bottom level) are then compared to how each program scored on each item within the specific assessment tool (see Figure 1).

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest level (top 20-25%)</td>
<td>A</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>Lowest level (bottom 20-25%)</td>
<td>C</td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td>Column Total</td>
<td>W</td>
<td>X</td>
<td>Grand Total</td>
</tr>
</tbody>
</table>

Once the data are sorted in the above matrix, the following formula (Figure 2) is used to determine if the standard is a key indicator or not by calculating its respective Phi coefficient. Please refer back to Figure 1 for the actual placement within the cells. The legend (Figure 3) below the formula shows how the cells are defined.

Figure 2 – Formula for Phi Coefficient

$$\phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}}$$

Figure 3 – Legend for the Cells within the Phi Coefficient

- A = High Group + Programs in Compliance on Specific Compliance Measure.
- B = High Group + Programs out of Compliance on Specific Compliance Measure.
- C = Low Group + Programs in Compliance on Specific Compliance Measure.
- D = Low Group + Programs out of Compliance on Specific Compliance Measure.
- W = Total Number of Programs in Compliance on Specific Compliance Measure.
- X = Total Number of Programs out of Compliance on Specific Compliance Measure.
- Y = Total Number of Programs in High Group.
- Z = Total Number of Programs in Low Group.
Once the data are run through the formula in Figure 2, the following chart (Figure 4) can be used to make the final determination of including or not including the item as a key indicator. Based upon the chart in Figure 4, it is best to have a Phi Coefficient approaching +1.00 however that is rarely attained with licensing data but has occurred in more normally distributed data. Continuing with the chart in Figure 5, if the Phi Coefficient is between +.25 and -.25, this indicates that the indicator is unpredictable in being able to predict overall compliance with the quality rating assessment tool. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance. This can occur with Phi Coefficients above +.25 but it becomes unlikely as we approach +1.00 although there is always the possibility that other standards/rules/regulations could be found out of compliance (this was demonstrated in a study conducted by the author. Another solution is to increase the number of key indicators to be reviewed but this will cut down on the efficiency which is desirable and the purpose of the key indicators.

The last possible outcome with the Phi Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the decision we want to make. The indicator would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously something we do not want to occur.

**Figure 4 – Thresholds for the Phi Coefficient**

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include</td>
</tr>
<tr>
<td>(.25) – (-.25)</td>
<td>Unpredictable</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>
RESULTS

Key indicators for child care homes (Please see the Appendix - Figure 7 for Phi Coefficients):

Section 406.8 General Requirements for Day Care Homes
a) The physical facilities of the home, both indoors and outdoors, shall meet the following requirements for safety to children.
   1) The home shall have a first aid kit consisting of adhesive bandages, scissors, thermometer, non-permeable gloves, Poison Control Center telephone number (1-800-222-1222 or 1-800-942-5969), sterile gauze pads, adhesive tape, tweezers and mild soap.

18) There shall be written plans for fire and tornado emergencies. Caregivers and assistants in the home shall be familiar with these plans.
   A) The fire evacuation plan shall identify the exits from each area used for child care and shall specify the evacuation route.
   B) The fire evacuation plan shall identify a safe assembly area outside of the home. It shall also identify a near-by indoor location for post-evacuation holding if needed.
   C) The fire evacuation plan shall require that the home be evacuated before calling the local emergency number 911.
   D) The written tornado plan shall specify what actions will be taken in the event of tornado or other severe weather warning, including designation of those areas of the home to be used as the safe spots.

23) The licensee shall inspect the home daily, prior to arrival of children, ensuring that escape routes are clear and that exit doors and exit windows are operable. A log of these daily inspections shall be maintained for at least one year, and shall be available for review. The log shall reflect, at minimum, the date and time of each inspection and the full name of the person who conducted it.

24) The licensee shall hold monthly fire inspections of the day care home.

Section 406.9 Characteristics and Qualifications of the Day Care Family
a) No individual may receive a license from the Department when the applicant, a member of the household age 13 and over, or any individual who has access to the children cared for in a day care home, or any employee of the day care home, has not authorized the background check required by 89 Ill. Adm. Code 385 (Background Checks) and been cleared in accordance with the requirements of Part 385.

   t) The caregivers shall complete 15 clock hours of in-service training per licensing year in accordance with the requirements in Appendix D of the rules.
      1) Such training may be derived from programs offered by any of the entities identified in Appendix D of the rules.
      2) Courses or workshops to meet this requirement include, but are not limited to, those listed in Appendix D of the rules.
      3) The records of the day care home shall document the training in which the caregiver has participated, and these records shall be available for review by the Department.
      4) Caregivers obtaining clock hours in excess of the required 15 clock hours per year may apply up to 5 clock hours to the next year's training requirements.

Section 406.12 Admission and Discharge Procedures
b) Prior to acceptance of a child for care,
   3) The caregiver shall require that the parent or guardian provide a certified copy of the child’s birth certificate. The caregiver:
A) Shall provide a written notice to the parent or guardian of a child to be enrolled for the first time that within 30 days after enrollment the parent or guardian shall provide a certified copy of the child’s birth certificate or other reliable proof of identity and age of the child.
   i) The caregiver shall promptly make a copy of the certified copy and return the original certified copy to the parent or guardian.
   ii) If a certified copy of the birth certificate is not available, the parent or guardian must submit a passport, visa or other governmental documentation as proof of the child’s identity and age and an affidavit or notarized letter explaining the inability to produce a certified copy of the birth certificate [325 ILCS 50/5].
   iii) The notice to parent or guardian shall also indicate that the caregiver is required by law to notify the Illinois State Police or local law enforcement agency if the parent or guardian fails to submit proof of the child’s identity within the 30 day time frame;

h) All day care homes shall have a written policy that explains the actions the provider will take if a parent or guardian does not retrieve, or arrange to have someone retrieve, his or her child at the designated, agreed upon time. The policy shall consist of the provider’s expectations, clearly presented to the parent or guardian, in the form of a written agreement that shall be signed by the parent or guardian, and shall include at least the following elements: The consequences of not picking up the child on time, including:
   - Amount of late fee, if any, and when those fees begin to accrue;
   - The degree of diligence the provider will use to reach emergency contacts, e.g., number of attempted phone calls to parents and emergency contacts, requests for police assistance in finding emergency contacts; and
   - Length of time the facility will keep the child beyond the pick-up time before contacting outside authorities, such as the child abuse hotline or police.

Emphasis on the importance of having up-to-date emergency contact numbers on file.
Acknowledgement of the provider’s responsibility for the child’s protection and well-being until the parent or outside authorities arrive.
A reminder to the day care provider that the child is not responsible for the situation. All discussions regarding these situations shall be with the parent or guardian, never the child.

Section 406.14 Health, Medical Care and Safety

c) A medical report, on forms prescribed by the Department, shall be on file for each child, on the first day of care, and shall be dated no earlier than 6 months prior to enrollment.
   1) The medical report shall be valid for 2 years, except that subsequent examinations for school-age children shall be in accordance with the requirements of Section 27.8-1 of the School Code [105 ILCS 5/27-8.1], provided copies of the exam are on file at the facility.
   2) If the child is in a high risk group, as determined by the examining physician, a tuberculin skin test by the Mantoux method and the results of that test shall be included in the initial examination for all children who have attained one year of age, or at the age of one year for children who are enrolled before their first birthday. The tuberculin skin test by the Mantoux method shall be repeated when the children in high-risk groups begin elementary and secondary school.
   3) The initial examination shall show that children from 6 months through 6 years of age have been screened for lead poisoning for children residing in an area defined as high risk by the Illinois Department of Public Health in its Lead Poisoning Prevention Code (77 Ill. Adm. Code 845) or that a lead risk assessment has been completed for children residing in an area defined as low risk by the Illinois Department of Public Health.
   4) The report shall indicate that the child has been immunized as required by the rules of the Illinois Department of Public Health for immunizations (77 Ill. Adm. Code 695). These required immunizations are poliomyelitis, measles, rubella, diphtheria, mumps, pertussis, tetanus, hepatitis B, haemophilus influenza B, and varicella (chickenpox) or provide proof of immunity according to requirements in Part 695.50 of the Department of Public Health.
Key indicators for Group Child Care Homes (Please see the Appendix - Figure 7 for Phi Coefficients):

Section 408.35 General Requirements for Group Day Care Home Family
f) The caregivers and all members of the household shall provide medical evidence that they are free of communicable disease that may be transmitted while providing child care; and, in the case of caregivers, that they are free of physical or mental conditions that could interfere with child care responsibilities. The medical report for the caregivers shall be valid for 3 years.

Section 408.45 Caregivers
f) The caregivers shall complete 15 clock hours of in-service training per licensing year in accordance with the requirements in Appendix G of the rules.
1) Such training may be derived from programs offered by any of the entities identified in Appendix G of the rules.
2) Courses or workshops to meet this requirement include, but are not limited to, those listed in Appendix G of the rules.

Section 408.60 Admission and Discharge Procedures
j) All group day care homes shall have a written policy that explains the actions the provider will take if a parent or guardian does not retrieve, or arrange to have someone retrieve, his or her child at the designated, agreed upon time. The policy shall consist of the provider's expectations, clearly presented to the parent or guardian in the form of a written agreement that shall be signed by the parent or guardian, and shall include at least the following elements:
1) The consequences of not picking up the children on time, including:
   A) Amount of late fee, if any, and when those fees begin to accrue;
   B) The degree of diligence the provider will use to reach emergency contacts, e.g., number of attempted phone calls to parents and emergency contacts, requests for police assistance in finding emergency contacts; and
   C) Length of time the facility will keep the child beyond the pick-up time before contacting outside authorities, such as the child abuse hotline or police.
2) Emphasis on the importance of having up-to-date emergency contact numbers on file.
3) Acknowledgement of the provider's responsibility for the child's protection and well-being until the parent or outside authorities arrive.
4) A reminder to staff that the child is not responsible for the situation. All discussions regarding these situations shall be with the parent or guardian, never with the child.

Section 408.70 Health, Medical Care and Safety
a) A medical report, on forms prescribed by the Department, shall be on file for each child, on the first day of care, and shall be dated no earlier than 6 months prior to enrollment.
1) The medical report shall be valid for 2 years, except that subsequent examinations for school-age children shall be in accordance with the requirements of Section 27-8.1 of the School Code [105 ILCS 5/27-8.1], provided copies of the exam are on file at the facility.
2) If the child is in a high risk group, as determined by the examining physician, a tuberculin skin test by the Mantoux method and the results of that test shall be included in the initial examination for all children who have attained one year of age, or at the age of one year for children who are enrolled before their first birthday. The tuberculin skin test by the Mantoux method shall be repeated when children in high risk groups begin elementary and secondary school.
3) The initial examination shall show that children from 6 months through 6 years of age have been screened for lead poisoning for children residing in an area defined as high risk by the Illinois Department of Public Health in its Lead Poisoning Prevention Code (77 Ill. Adm. Code 845) or that a lead risk assessment has been completed for children residing in an area defined as low risk by the Illinois Department of Public Health.
4) The report shall indicate that the child has been immunized as required by the rules of the Illinois Department of Public Health for immunizations (77 Ill. Adm. Code 695). These required immunizations are poliomyelitis, measles, rubella, diphtheria, mumps, pertussis, tetanus, hepatitis B, haemophilus influenza B, and varicella (chickenpox) or provide proof of immunity according to requirements in Part 695.50 of the Department of Public Health.
Section 408.120  Records and Reports
a) A facility shall maintain a record file on the children enrolled.
1) A written application for admission of each child shall be on file with the signature of the parent or guardian.
Key indicators for Child Care Centers (Please see the Appendix-Figure 7 for Phi Coefficients):

Section 407.100 General Requirements for Personnel
f) Staff shall have physical re-examinations every two years and whenever communicable disease or illness is suspected.

Section 407.120 Personnel Records
a) A confidential file shall be maintained on each staff person and contain at least the following information:
1) A copy of a form prescribed by the Department which contains information on persons employed in the day care center;
3) Three written character references, verified by the day care center;
4) Proof of educational achievement as required for the individual's position. Foreign credentials require additional documentation providing a statement of the equivalency in the U.S. educational system;

Section 407.250 Enrollment and Discharge Procedures
d) The facility shall distribute a summary of the licensing standards, provided by the Department, to the parents or guardian of each child at the time that the child is accepted for care in the facility. In addition, consumer information materials provided by the Department including, but not limited to, information on reporting and prevention of child abuse and neglect and preventing and reporting communicable disease shall be distributed to the parents or guardian or each child cared for when designated for such distribution by the Department.

Section 407.260 Daily Arrival and Departure of Children
f) All day care centers shall have a written policy that explains to parents and staff the actions the center will take if a parent or guardian does not pick up, or arrange to have someone pick up, his or her child at the designated, agreed upon time. The policy shall consist of the provider's expectations clearly presented to the parent or guardian in the form of a written agreement that shall be signed by the parent or guardian and shall include at least the following elements:
1) The consequences of not picking up children on time shall be precisely communicated to parents, for example:
   A) Amount of late fee, if any, and when those fees begin to accrue.
   B) The degree of diligence the provider will use to reach emergency contacts, e.g., number of attempted phone calls to parents and emergency contacts, requests for police assistance in finding emergency contacts, and so forth.
   C) Length of time the facility will keep the child beyond the pick-up time before contacting outside authorities, such as, the child abuse hotline, police, and so forth.
2) Emphasis on the importance of having up-to-date emergency contact numbers on file.
3) Acknowledgement of the provider's responsibility for the child's protection and well-being until the parent or outside authorities arrive.
4) A policy that staff shall not hold the child responsible for the situation and that discussion of this issue will only be with the parent or guardian and never with the child.

Section 407.270 Guidance and Discipline
a) The day care center shall develop a guidance and discipline policy for staff use that is also provided to parents. Staff shall sign the guidance and discipline policy at the time of employment and parents shall sign the policy when their child is enrolled. The policy shall include:
1) A statement of the center's philosophy regarding guidance and discipline;
2) Information on how discipline will be implemented by staff;
3) Information on how parents will be involved in the guidance and discipline process;
4) Information on how children will be involved in the guidance and discipline process; and
5) Written procedures for termination of a child's enrollment in the day care center because of disciplinary issues.

Section 407.310 Health Requirements for Children
a) A medical report on forms prescribed by the Department shall be on file for each child.
1) The initial medical report shall be dated less than 6 months prior to enrollment of infants, toddlers and preschool children. For school-age children, a copy of the most recent regularly scheduled school physical may be submitted
(even if more than 6 months old) or the day care center may require a more recent medical report by its own enrollment policy. If a health problem is suspected, the day care center may require additional documentation of the child's health status.

Section 407.380 Equipment and Materials
b) Such equipment and materials for infants, toddlers and pre-school children shall be provided in the quantity and variety specified in Appendix A: Equipment for Infants and Toddlers, Appendix B: Equipment for Preschool Children and Appendix C: Equipment for School-Age Children of the Rules.
For additional information regarding this Report, please contact:
Richard Fiene, Ph.D., Director/President
Research Institute for Key Indicators (RIKI)
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
717-944-5868 Phone and Fax
http://RIKInstitute.wikispaces.com
Appendix – Figure 5

DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA®) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model

\[ CI \times PQ \Rightarrow RA + KI \Rightarrow DM + PD \Rightarrow CO \]

Definitions of Key Elements:

- **PC** = Program Compliance/Licensing (Health and Safety) *(Caring for Our Children)*
- **PQ** = QRIS/Accreditation/Caregiver/Child Interactions/Classroom Environment Quality *(ERS/CLASS/PAS/BAS)*
- **RA** = Risk Assessment, (High Risk Rules) *(Stepping Stones)*
- **KI** = Key Indicators (Predictor Rules) *(13 Key Indicators of Quality Child Care)*
- **DM** = Differential Monitoring (How often to visit and what to review)
- **PD** = Professional Development/Technical Assistance/Training *(Not pictured but part of Model)*
- **CO** = Child Outcomes *(Not pictured but part of Model)*
Appendix – Figure 6 - Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

All Licensing Rules – Full Compliance Reviews

Differential Monitoring

How Often to Visit? What is Reviewed?

Frequency

More Often

Less Often

Abbreviated Tool

Risk Assessment Weights

Key Indicators Predictors
### Appendix -- Figure 7 - Phi Coefficients for the Specific Key Indicators

#### Family Child Care Homes:

<table>
<thead>
<tr>
<th>Rule Numbers</th>
<th>Phi</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>406.8a1</td>
<td>.34</td>
<td>First Aid Kit</td>
</tr>
<tr>
<td>406.8a18</td>
<td>.38</td>
<td>Emergency Plan</td>
</tr>
<tr>
<td>406.8a23</td>
<td>.36</td>
<td>Fire Inspection</td>
</tr>
<tr>
<td>406.8a24</td>
<td>.35</td>
<td>Log of Home Inspections</td>
</tr>
<tr>
<td>406.9a</td>
<td>.34</td>
<td>Background Checks</td>
</tr>
<tr>
<td>406.9t</td>
<td>.38</td>
<td>Caregiver Training</td>
</tr>
<tr>
<td>406.12b3</td>
<td>.34</td>
<td>Birth Certificate</td>
</tr>
<tr>
<td>40612h</td>
<td>.36</td>
<td>Agreement regarding Pick Up</td>
</tr>
<tr>
<td>406.14c2</td>
<td>.41</td>
<td>TB Test</td>
</tr>
<tr>
<td>406.14c3</td>
<td>.53</td>
<td>Lead Poisoning Screening</td>
</tr>
<tr>
<td>406.14c4</td>
<td>.34</td>
<td>Immunizations</td>
</tr>
</tbody>
</table>

#### Group Child Care Homes:

<table>
<thead>
<tr>
<th>Rule Numbers</th>
<th>Phi</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>408.35f</td>
<td>.28</td>
<td>Communicable Diseases</td>
</tr>
<tr>
<td>408.45f</td>
<td>.31</td>
<td>Caregiver Training</td>
</tr>
<tr>
<td>408.60j</td>
<td>.33</td>
<td>Agreement Pick Up Policy</td>
</tr>
<tr>
<td>408.70a1</td>
<td>.29</td>
<td>Medical Records</td>
</tr>
<tr>
<td>408.70a2</td>
<td>.55</td>
<td>TB Test</td>
</tr>
<tr>
<td>408.70a3</td>
<td>.51</td>
<td>Lead Poisoning Screening</td>
</tr>
<tr>
<td>408.70a4</td>
<td>.35</td>
<td>Immunizations</td>
</tr>
<tr>
<td>408.120a1</td>
<td>.37</td>
<td>Written Application Admission for Each Child</td>
</tr>
</tbody>
</table>

#### Child Care Centers:

<table>
<thead>
<tr>
<th>Rule Numbers</th>
<th>Phi</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>407.100f</td>
<td>.35</td>
<td>Staff Physical</td>
</tr>
<tr>
<td>407.120a1</td>
<td>.32</td>
<td>CFS-508 Form</td>
</tr>
<tr>
<td>407.120a3</td>
<td>.41</td>
<td>Three Written Character References</td>
</tr>
<tr>
<td>407.120a4</td>
<td>.34</td>
<td>Proof of Educational Achievement</td>
</tr>
<tr>
<td>407.250d</td>
<td>.34</td>
<td>Written Standards Given to Parents</td>
</tr>
<tr>
<td>407.260f</td>
<td>.32</td>
<td>Pick Up Policy</td>
</tr>
<tr>
<td>407.270a</td>
<td>.32</td>
<td>Discipline Policy</td>
</tr>
<tr>
<td>407.310a</td>
<td>.44</td>
<td>Medical Report for Each Child</td>
</tr>
<tr>
<td>407.380b</td>
<td>.34</td>
<td>Equipment Meets Standard Requirements</td>
</tr>
</tbody>
</table>
Wisconsin Department of Children and Youth Services Program Monitoring Options Blueprint Report

Richard Fiene, Ph.D.

May 15, 2014

ABSTRACT

This report will provide a blueprint for consideration by Wisconsin’s Office of Children and Youth Services regarding options for their program monitoring system. The report will be organized into the following major headings: an introduction to program monitoring; how key indicators and risk assessment fit into the larger program monitoring of human services; how key indicators and risk assessment could be applied to Wisconsin’s system in particular; the technical aspects of differential monitoring, risk assessment and key indicator methodology, the sample to be drawn from the population, a timeline for this developmental effort; and potential cost savings from the approach. Many of the examples drawn are from the child care/early care and education field rather than the child welfare/child residential field because most of the best examples are occurring in child care and not child welfare at this point in time. Hopefully, with this blueprint is implemented in children and youth services, we can begin to change this fact.

INTRODUCTION

An effective and efficient program monitoring system is a goal of every state human service agency in the USA. This has been an issue in the human services for over the past half century as states grapple with increasing caseload sizes with shrinking resources. This report will provide an overview to the topic and several options that the State of Wisconsin can begin to explore related the program monitoring of children and youth services. The Risk Assessment, Key Indicator, and Differential Program Monitoring Methodologies were developed to help streamline the program monitoring of early care and education programs. It was first applied in child care licensing (Fiene & Nixon, 1985) but has been used in many other service types, such as: Head Start Performance Standards (Fiene, 2013a), National Accreditation (Fiene, 1996), and child and adult residential programs (Kroh & Melusky, 2010). The methodologies are based
upon statistical protocols that have been developed in the tests and measurements literature in which an abbreviated set of items is used to statistically predict as if the full test was applied. This methodology has been used in regulatory analysis and more recently has been proposed for use in Quality Rating and Improvement Systems (QRIS) (Fiene, 2013b). In reviewing the various states and the research literature, one state did not come to the surface with all the components in place for child welfare/child residential services, therefore a preponderance of examples drawn from the child care/early care and education field are used throughout the report. However, there are many similarities obviously from child care to child welfare with the most obvious being the protection of children and “to do no harm” as the ultimate outcome of services.

DIFFERENTIAL PROGRAM MONITORING

Risk Assessment and Key Indicators are important components of differential program monitoring which employs an abbreviated review rather than a comprehensive or full review of a program. It is one of several key elements that have been identified in the research literature to help improve the cost effectiveness and efficiency of the program monitoring of early care and education programs (Fiene, 2013b, c) (See the Appendix for two graphics that depict the key elements). A recent addition to differential monitoring are QRIS – Quality Rating and Improvement Systems. Key indicators have a long history of development within the licensing literature (Fiene & Kroh, 2000) but have not had a long history in child and adult residential services. This proposed blueprint is to assist Wisconsin to develop a fully functional differential program monitoring, risk assessment, and key indicator approach to their licensing system and then determine the cost and resources needed in implementing this approach.

The graphics in the Appendix depict the critical key elements of a differential program monitoring approach. In the first graphic program compliance/licensing is generally a state’s health and safety rules/regulations. The program quality key element for children and youth services would generally be represented by the national standards, such as the Child Welfare League of America’s Standards. The key indicator element is represented by the state’s statistical predictor rules/regulations drawn from their comprehensive set of rules/regulations. The last key element to be addressed in this report is the risk assessment key element in which these are the high risk rules/regulations that place children at greatest risk of mortality or morbidity. All these key elements will be addressed in this report in greater detail outlining the technical aspects of each. The second graphic in the Appendix – Graphic 2 depicts the relationship between licensing rules, compliance reviews, differential monitoring, abbreviated tools, risk assessment and key indicators. As one can see from this graphic it demonstrates the inter-relationships amongst all the program monitoring components.
KEY INDICATORS APPLIED TO WISCONSIN'S CHILDREN AND YOUTH LICENSING SYSTEM

Before beginning the description of each of the key elements it is important to note that there are some significant challenges because of the psychometric properties of licensing data such as the severe skewness and kurtosis present in state licensing data systems. These challenges will be addressed later in this blueprint in how to deal with skewness and kurtosis.

As a footnote, the risk assessment and key indicators can eventually be tied to the professional development/training/technical assistance system to link resources to specific needs of the programs. It also has the capability of tying them to specific child outcomes at some point in the future. This would be accomplished in the full implementation of the Differential Monitoring Logic Model and Algorithm (DMLMA©) as depicted in the Appendix – Graphic 1.

TECHNICAL ASPECTS OF THE KEY INDICATOR METHODOLOGY

This section provides the technical and statistical aspects of the key indicator methodology. It will provide the roadmap in taking the Wisconsin licensing data base through the necessary steps to generating the respective key indicators.

One of the first steps is to sort the data into high and low groups, generally the highest and lowest ratings can be used for this sorting. In very large states this is done on a sampling basis which will be described later in the blueprint. Frequency data will be obtained on those programs in the top level (usually top 20-25%) and the bottom level (usually the bottom 20-25%). The middle levels are not used for the purposes of these analyses. These two groups (top level & the bottom level) are then compared to how each program scored on each item within the specific assessment tool (see Figure 1). An example is provided in Figure 2 from a previous study conducted by the author (see Figure 2).

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highest level (top 20-25%)</strong></td>
<td>A</td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Lowest level (bottom 20-25%)</strong></td>
<td>C</td>
<td>D</td>
<td>Z</td>
</tr>
<tr>
<td><strong>Column Total</strong></td>
<td>W</td>
<td>X</td>
<td><strong>Grand Total</strong></td>
</tr>
</tbody>
</table>
Figure 2 depicts that all programs that were in the top 25% were also in the highest rating while the bottom 25% were also in the lowest rating. The data depicted in Figure 2 are taken from studies completed in Pennsylvania in 2002 (Fiene, et al) and 2006 (Barnard, Smith, Fiene & Swanson) in which their quality rating and improvement system, Keystone STARS, was validated. The reason for selecting this particular item from the ECERS – Early Childhood Environment Rating Scale is that it demonstrates a perfect phi coefficient in discriminating between the highest level and the lowest level. Most, if not all, of the licensing items that will attain the threshold levels to become key indicators will not approach this phi coefficient.

<table>
<thead>
<tr>
<th>Figure 2 – Pa. Study (Fiene, et al, 2002).</th>
<th>Providers In Compliance or Top 25%</th>
<th>Programs Out Of Compliance or Bottom 25%</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Star level in Pa.</td>
<td>117</td>
<td>0</td>
<td>117</td>
</tr>
<tr>
<td>Lowest Star level in Pa.</td>
<td>0</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Column Total</td>
<td>117</td>
<td>35</td>
<td>152</td>
</tr>
</tbody>
</table>

Once the data are sorted in the above matrix, the following formula (Figure 3) is used to determine if Item 16 is a key indicator or not by calculating its respective Phi coefficient. Please refer back to Figure 1 for the actual placement within the cells and Figure 2 for the data within the cells. The legend (Figure 4) below the formula shows how the cells are defined.

Figure 3 – Formula for Phi Coefficient

\[
\phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}}
\]

Figure 4 – Legend for the Cells within the Phi Coefficient

- **A** = High Group + Programs in Compliance on Specific Compliance Measure.
- **B** = High Group + Programs out of Compliance on Specific Compliance Measure.
- **C** = Low Group + Programs in Compliance on Specific Compliance Measure.
- **D** = Low Group + Programs out of Compliance on Specific Compliance Measure.
- **W** = Total Number of Programs in Compliance on Specific Compliance Measure.
- **X** = Total Number of Programs out of Compliance on Specific Compliance Measure.
- **Y** = Total Number of Programs in High Group.
- **Z** = Total Number of Programs in Low Group.
Once the data are run through the formula in Figure 3, the following chart (Figure 5) can be used to make the final determination of including or not including the item as a key indicator. Based upon the chart in Figure 5, it is best to have a Phi Coefficient approaching +1.00 if we are dealing with normally distributed data.

Continuing with the chart in Figure 5, if the Phi Coefficient is between +.25 and -.25, this indicates that the indicator is unpredictable in being able to predict overall compliance with the quality rating assessment tool. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance. This can occur with Phi Coefficients above +.25 but it becomes unlikely as we approach +1.00 although there is always the possibility that other standards/rules/regulations could be found out of compliance (this was demonstrated in a study conducted by the author (Fiene, 2013c). Another solution is to increase the number of key indicators to be reviewed but this will cut down on the efficiency which is desirable and the purpose of the key indicators.

The last possible outcome with the Phi Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the decision we want to make. The indicator would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously something we do not want to occur.

Figure 5 – Thresholds for the Phi Coefficient (Fiene & Nixon, 1983, 1985)

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include</td>
</tr>
<tr>
<td>(+.25) – (-.25)</td>
<td>Unpredictable</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>

The key indicators should then only be used with those programs who have attained the highest rating. It is not intended for those programs that have attained lower ratings. However, even with those programs that have attained the highest rating, every 3-5 years a full, comprehensive review using the full set of rules/standards for licensing should occur (see Figure 6 for a graphical depiction). It is intended that a re-validation of the key indicators occur on a periodic basis to make certain that the key indicators have not changed because of differences in compliance history. This is an important and necessary step for the state to engage in to
ascertain the overall validity and reliability of the assessment system. Also there should not have been any major changes in the program while the key indicators are being administered, such as the director/administrator leaving or a large percentage of staff leaving or caseloads increasing significantly, or a change in the licensing status of the program.

**Figure 6 - Proposed DMLMA System with Key Indicators (KI)**

*Use of Wisconsin Key Indicators (WKI) for Licensing with a Full Review every 4th Year*

![Diagram of DMLMA System](image)

**TECHNICAL ASPECTS OF THE RISK ASSESSMENT METHODOLOGY**

The risk assessment methodology is very different from the key indicator methodology in that compliance history data are not utilized but rather a best practice ranking according to risk is used to determine which rules become core rules which have the greatest likelihood to place children at significant risk of morbidity or mortality. This is done by having a group of experts rank order all the rules on a Likert Scale from low risk to high risk of mortality or morbidity that non-compliance with the rule places children at. This is generally done on a 1-10 scale with 1 = low risk; 5 = medium risk; and 10 = high risk (see Figure 6A). The experts selected include but are not limited to licensing staff, policy makers, researchers, providers, advocacy groups, parents, and other significant stakeholders who will be impacted by the weighting of the rules.

**Figure 6A – Example of a Likert Scale for Measuring Risk to Children**

<table>
<thead>
<tr>
<th>Low Risk</th>
<th>Medium Risk</th>
<th>High Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Once the data are collected from all the experts, it is averaged for each rule to determine its relative rank in comparison to all the other rules. A significantly high threshold or cut off point is determined so that no more than 5-10% of the rules become core rules. These core rules can then be used in a differential monitoring approach (to be described more fully in the next section) and/or with the key indicators to complete abbreviated reviews of child welfare programs. It is recommended that such a practice of using both core rules and key indicators be used together
because than the state has the benefits of both methodologies in measuring risk and being able to statistically predict overall compliance with a very short list of rules.

TECHNICAL ASPECTS DIFFERENTIAL MONITORING METHODOLOGY

There are a couple of other key technical aspects that need to be in place for a differential monitoring system to work. The Differential Monitoring Logic Model and Algorithm (DMLMA©)(see the Appendix) is a 4th generational Early Childhood Program Quality Indicator Model4 (ECPQIM4©) in which the major monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With this new model, it is now possible to compare results obtained from licensing systems, quality assurance systems, risk assessment systems, key indicator systems, technical assistance, and child protection outcome systems. The various approaches to validation are interposed within this model and the specific expected correlational thresholds that should be observed amongst the key elements of the model are suggested (see Figure 6B).

Figure 6B – Inter-Correlational Threshold Matrix

<table>
<thead>
<tr>
<th></th>
<th>PQ</th>
<th>RA</th>
<th>KI</th>
<th>DM</th>
<th>PD</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.3</td>
<td>0.5</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>PQ</td>
<td></td>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>KI</td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

Key Elements (see the Appendix): CI = state or federal standards, usually rules or regulations. PQ = CWLA Standards or a Quality Assurance System. RA = risk assessment tools/systems in
which only the most critical rules/standards are measured. KI = key indicators in which only predictor rules/standards are measured. DM = differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol. PD = technical assistance/training and/or professional development system which provides targeted assistance to the program based upon the DM results. CO = child outcomes which assesses how well the children are protected which is the ultimate goal of the system.

Once the above key elements are in place, it is then possible to look at the relationships amongst them to determine if the system is operating as it was intended. This is done through a validation of the overall system and assessing the inter-correlations (Table 6B) to determine that the DM system is improving the overall protection of the children it serves.

Wisconsin could use the following plan to implement the above approach:

STATE AGENCY PLAN (These Steps can be viewed as an overall plan as outlined in Zellman & Fiene (2012):

The first step in utilizing the DMLMA for a state is to take a close look at its Comprehensive Licensing Tool (CI) that it uses to collect violation data on all rules with all facilities in its respective state. If the state does not utilize a tool or checklist or does not review all violation data than it needs to consider these changes because the DMLMA is based upon an Instrument Based Program Monitoring System (IPM) which utilizes tools/checklists to collect data on all rules.

The second step for the state is to compare their state’s rules with the National Standards (such as the CWLA National Standards for Best Practices) to determine the overlap and coverage between the two. This is the first approach to validation which involves Standards review (Zellman & Fiene, 2012).

The third step for the state is to compare the results from the CI with the RA tools. This step is the second approach to validation which involves Measures (Zellman & Fiene, 2012). The correlation between CI and RA should be at the .50 level or higher (.50+)(see Figure 6B).

The fourth step is for the state to generate a Key Indicator (KI) tool from the CI data base. Please see Fiene & Nixon (1985) and Fiene & Kroh (2000) for a detailed explanation of the methodology for generating a KI tool. This step is also part of the second approach to validation which involves Measures. The correlation between the CI and KI should be very high (.70+) because the KI is a subset of predictor rules taken from the CI data base.

The fifth step for the state is to use the RA and KI tools together to determine overall compliance of facilities and how often and which rules will be monitored for future visits. This is the basic
component of a Differential Monitoring (DM) approach and continues the second approach to validation (Measures). Also, this step should drive decisions within the technical assistance/training/professional development (PD) system in what resources are allocated to a particular facility. It would be expected that moderate correlations (.50+) would be found amongst RA, KI, DM, and PD.

The sixth and final step for the state is to compare the results from the various monitoring tools (CI, PQ, RA, KI) with any child development outcome (CO) data they collect. This is a relatively new area and few, if any, states at this point have this capability on a large scale. This step is the fourth approach to validation which involves Outcomes (Zellman & Fiene, 2012). The correlations between CI, PQ, RA, KI and CO will be on the lower end (.30+) because there are so many other variables that impact the child other than child welfare services.

The last step is to present a logic model which depicts how a differential monitoring system could potentially be actually used in Wisconsin (see Figure 6C).

**Figure 6C – Logic Model for Compliance Decisions**

![Logic Model](image-url)

**Compliance Decisions:**

*Core Indicators = Core Rules + Key Indicators – this becomes a screening tool to determine if a program receives a AV or FV visit.*

*Core Indicators (100%) = the next visit is a Abbreviated Visit. Every 3-4 years a Full Licensing Visit is conducted.*

*Core Indicators (not 100%) = The next visit is a Full Licensing Visit where all rules are reviewed.*

*Compliance = 96%+ with all rules which indicates substantial to full compliance with all rules and 100% with Core Indicators. The next visit is an Abbreviated Visit.*

*Non-compliance = less than 96% with all rules which indicates lower compliance with all rules. The next visit is a Full Visit Study.*
SAMPLE

Generally a sample is drawn from the population of early care and education facilities in respective states. Wisconsin will not be any different because of the size of the overall child welfare program. A random sample will be selected that represents the state population of child welfare programs. This will be determined by the number of programs, how the programs are distributed throughout the state, the size of the programs, the type of programs, etc… This will need to be determined once the actual implementation of this blueprint report is started. The author of this report can assist Wisconsin staff in how best to select the sample of programs.

TIMELINE

As soon as all the Wisconsin child welfare/child residential programs have gone through their assessment phase, it will be possible to do the calculations to determine the Phi Coefficients and generate the key indicators. I am guessing that this should not take any longer than 1 year but could be completed in a much shorter period of time if the assessments on individual programs could be moved up (see Figure 7). The analytical phase should take no longer than a month with an additional month to write up the report. A face to face presentation of the analyses could be done after these two months.

The timeline presented in Figure 7 can be adjusted to the specific needs for the Wisconsin system. The timeline is based upon previous projects and the average time to generate risk assessment core rules and key indicators. Another consideration or task is the development of the policies and procedures to be developed and implemented regarding the use of key indicators. This was not specifically listed on the timeline because it is something that is generally developed throughout the project with feedback from all the stakeholders who will be impacted by the use of this new approach to assessment and monitoring.

Figure 7 - WISCONSIN DMLMA PROJECT TIMELINE

<table>
<thead>
<tr>
<th>TASK</th>
<th>MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect Data</td>
<td>M1-M3</td>
</tr>
<tr>
<td>Sort Data</td>
<td>M2-3</td>
</tr>
<tr>
<td>Run Analyses</td>
<td>M3-5</td>
</tr>
<tr>
<td>Generate KI/RA</td>
<td>M6</td>
</tr>
<tr>
<td>Train on KI/RA</td>
<td>M6-7</td>
</tr>
<tr>
<td>KI/RA Reliable Implementation</td>
<td>M7-9</td>
</tr>
<tr>
<td>Implementation</td>
<td>M10-12</td>
</tr>
</tbody>
</table>
Legend:

KI – Key Indicators
RA – Risk Assessment

Collect Data – identify participant programs via sampling for KI and the stakeholders for RA.
Sort Data – KI - the individual programs are sorted into high and low groups representing the top 25% and the bottom 25% of programs as they have scored on the respective rules/standards.
Run Analyses – KI - each individual item within each of the assessment tools for every program will be compared to the sorting process of the high and low groups. RA – aggregate data into means for each rule, rank order the rules.
Generate KI/RA – a 2 x 2 matrix is constructed and the key indicators (KI) are generated from this matrix through the use of a phi coefficient. A final report will be delivered to Wisconsin executive staff for both KI and RA core indicator rules.
Training on KI/RA – all staff who will be using the KI/RA will be trained on its use.
KI/RA Reliability – reliability will be established by having two staff go out together and administer the key indicators separately and comparing their results.
Implementation – once reliability has been established, full implementation will begin.

COST SAVINGS

Again based upon previous studies most recently completed in California in 2010 (http://www.myccl.ca.gov/res/docs/12022010HandoutStakeholderMeeting.pdf), time savings of 50% have been attained by using a key indicator or abbreviated tool in completing assessments. It only makes sense that if an assessment can be completed in one hour rather than 2 – 4 hours that a state will see time savings. It is being assumed that equivalent savings should also be the case with Wisconsin’s licensing system although this cannot be made certain until the new key indicator or abbreviated tool is actually used for a period of time. Once the new key indicators are used for several months, comparisons could be made to when the full assessments were done.

CONCLUSION, OPTIONS, AND RECOMMENDATIONS

This blueprint report has given the basic empirical parameters to develop a differential monitoring, risk assessment, and key indicator approach to Wisconsin’s Children and Youth Licensing system⁴. By following this blueprint Wisconsin staff should be able to fully implement the approach. Wisconsin staff would also need to determine if they have the internal capability for the development of the key indicators or if there will be the need to outsource certain aspects of the development. This will be an important consideration as Wisconsin moves forward with this project. I have provided two options for your consideration in moving forward.

Option 1 – Development of System Internally:

This would require either information systems or research & evaluation staff to analyze the data, generate core key indicator rules, and training of staff. I could provide the necessary consulting services to help the staff work through the methodology. This would probably require at least one face to face meeting with regular monthly conference calls between myself and staff. Discussions of the formatting of data and the types of analyses would be discussed and
demonstrated. The overall cost to develop the system internally with NARA support would be approximately $100,000.

**Option 2 – Development of System Externally:**

In this option I could do all the methodological work demonstrating how I would need the data sent to me, the analytical work in generating core key indicator rules, a report detailing the methodology and results. The only thing that Wisconsin staff would need to do is get the data to me, all other aspects of what is delineated in the timeline in Figure 7 would be completed by me. This would probably require several face to face trips to explain the process, the results, and do training of staff. Once everything was in place, Wisconsin staff would have a fully implemented system. The overall cost to develop the system externally with NARA support would be approximately $300,000.

Whatever option is selected the following *recommendations* are provided if Wisconsin staff want to develop a program monitoring system based upon empirical data:

1. Wisconsin should move forward with enhancing their differential monitoring approach in order to institute potential cost savings and reallocation of resources based upon those cost savings.

2. Develop and implement a key indicator approach based upon the methodology described in this blueprint.

3. Develop and implement a risk assessment approach based upon the methodology described in this blueprint.

4. A staff caseload analysis should be completed based upon *NARA’s Licensing Workload Assessment* in order to determine the exact number of additional staff needed to fully implement a Differential Monitoring Approach.
Notes:

1. The reason for pointing out the need to have a higher Phi Coefficient than what has been reported previously (Fiene & Nixon, 1983, 1985) is the fact that the dichotomization of data should only be used with skewed data and not normally distributed data because it will accentuate differences. However, since the purpose of the dichotomization of data is only for sorting into a high and low group, it would appear to be acceptable for this purpose (MacCallum, et al., 2002. On the practice of dichotomization of quantitative variables, Psychological Methods, 7, 1, 19-40.).

2. These results would show an increase in cells B and C in Figure 1 which is undesirable; it should always be the case where A + D > B + C for key indicators to maintain their predictive validity. The distinction between making decisions with skewed (Licensing) as versus normally distributed (ERS) data is an important one because there is a greater likelihood with skewed data of introducing less than optimal programs into the high group when sorting programmatic data into high and low groups. This then makes it more difficult to identify the best programs. However, because of the distribution with skewed data the same cannot be said with the low group in which case it is relatively easy to identify the problem programs. This is not as much of a concern when the data are more normally distributed in which it is relatively easy to identify both the optimal and problem programs. This is an excellent example of the need of weighting of standards in order to increase the normal distribution of the data.

3. It is important to note that many of the examples are drawn from the child care research literature and not from the child welfare research literature. The reason for this is most of the empirical basis for the development of these methodologies was completed in child care over the past 40 years. It is important for the reader of this report to keep this in mind and to make the necessary translations to the child welfare literature research base. For example, when I describe the national health and safety standards in child care, the reader should be thinking of the CWLA national standards for the various child welfare service types. QRIS systems can translate to child welfare systems that locally have been built upon generic licensing systems. The DMLMA model is a generic model for all human services and not only for child care, so the reader should be able to make the translation from child care to child welfare.

4. There are two publications that are more pertinent to children & youth services and child welfare that I wrote back in the 1980’s the Wisconsin staff may be interested in (Fiene & McDonald, (1987), Instrument Based Program Monitoring and Indicator Checklist for Child Welfare, and Fiene (1981), Conceptual Framework for Program Monitoring).

For additional information regarding this Blueprints Report, please contact:
Richard Fiene, Ph.D., Director/President
Research Institute for Key Indicators (RIKI)
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
717-944-5868 Phone and Fax
http://RIKInstitute.wikispaces.com
REFERENCES AND ADDITIONAL RELATED READINGS REGARDING DIFFERENTIAL MONITORING, RISK ASSESSMENT, AND KEY INDICATOR METHODOLOGIES:


**Appendix – Graphic 1**

**DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA®) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model**

\[ CI \times PQ \Rightarrow RA + KI \Rightarrow DM + PD \Rightarrow CO \]

**Definitions of Key Elements:**

- **PC** = Program Compliance/Licensing (Health and Safety, Protections for Children)
- **PQ** = QRIS/Accreditation/Caregiver/Child Interactions
- **RA** = Risk Assessment, (High Risk Rules)
- **KI** = Key Indicators (Predictor Rules)
- **DM** = Differential Monitoring (How often to visit and what to review)
- **PD** = Professional Development/Technical Assistance/Training (Not pictured but part of Model)
- **CO** = Child Outcomes (Not pictured but part of Model)
Appendix – Graphic 2 - Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

All Licensing Rules – Full Compliance Reviews

Differential Monitoring

How Often to Visit? What is Reviewed?

Frequency

More Often

Less Often

Abbreviated Tool

Risk Assessment Weights

Key Indicators Predictors
The purpose of this report is to present to the Office of Head Start (OHS) Key Indicators of their Head Start Performance Standards (HSPS) that have the ability to statistically predict substantial compliance with all Compliance Measures and ultimately the majority of HSPS’s. The analytical and methodological basis of this approach is based upon a *Differential Monitoring Logic Model and Algorithm (DMLMA®)* (Fiene, 2012) (see Appendix 3). The DMLMA® is the 4th generation of an Early Childhood Program Quality Indicator Model (ECPQIM)(Fiene & Nixon, 1985; Griffin & Fiene, 1995; Fiene & Kroh, 2000). Only a portion of the DMLMA® model was utilized in this report which focused on key indicators, risk assessment, and program quality.

Definitions:

**Risk Assessment (RA)** - a differential monitoring approach that employs using only those rules, standards, or regulations that place children at greatest risk of mortality or morbidity if violations/citations occur with the specific rule, standard, or regulation.

**Key Indicators (KI)** - a differential monitoring approach that employs using only those rules, standards, or regulations that statistically predict overall compliance with all the rules, standards, or regulations. In other words, if a program is 100% in compliance with the Key Indicators the program will also be in substantial to full compliance with all rules, standards, or regulations. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators the program will also have other areas of non-compliance with all the rules, standards, or regulations.

**Differential Monitoring (DM)** - this is a relatively new approach to determining the number of visits made to programs and what rules, standards, or regulations are reviewed during these visits. There are two measurement tools that drive differential monitoring, one is Weighted Risk Assessment tools and the other is Key Indicator checklists. Weighted Risk Assessments determine how often a program will be visited while Key Indicator checklists determine what rules, standards, or regulations will be reviewed in the program. Differential monitoring is a very powerful approach when Risk Assessment is combined with Key Indicators because a program is reviewed by the most critical rules, standards, or regulations and the most predictive rules, standards, or regulations. See Appendix 3 which presents a Logic Model & Algorithm for Differential Monitoring (DMLMA®)(Fiene, 2012).

**Program Quality (PQ)** - for the purposes of this study this was measured via the CLASS – Classroom Assessment Scoring System. The CLASS has three sub-scales (ES = Emotional Support, CO = Classroom Organization, and IS = Instructional Support). The CLASS is a tool that is identified in the research literature as measuring classroom quality similar to the ERS tools.
Early Childhood Program Quality Indicator Model (ECPQIM) — these are models that employ a key indicator or dashboard approach to program monitoring. Major program monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With these models, it is possible to compare results obtained from licensing systems, quality rating and improvement systems (QRIS), risk assessment systems, key indicator systems, technical assistance, and child development/early learning outcome systems. The various approaches to validation are interposed within this model and the specific expected correlational thresholds that should be observed amongst the key elements of the model are suggested. Key Elements of the model are the following (see Appendix 3 for details): CI = state or federal standards, usually rules or regulations that measure health and safety - Caring for Our Children or Head Start Performance Standards will be applicable here. PQ = Quality Rating and Improvement Systems (QRIS) standards at the state level; ERS (ECERS, ITERS, FDCRS), CLASS, or CDPES (Fiene & Nixon, 1985). RA = risk assessment tools/systems in which only the most critical rules/standards are measured. Stepping Stones is an example of this approach. KI = key indicators in which only predictor rules/standards are measured. The Thirteen Indicators of Quality Child Care is an example of this approach. DM = differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol. PD = technical assistance/training and/or professional development system which provides targeted assistance to the program based upon the DM results. CO = child outcomes which assesses how well the children are developing which is the ultimate goal of the system.

The organization of this report is as follows:

1) The first section will provide an overall analysis the Head Start (HS), Early Head Start (EHS), and Head Start/Early Head Start (HS/EHS) programs; 1,4

2) The second section will provide analyses of the various content areas (CA) within the HSPS; 4

3) The third section will provide analyses of the relationship between the HSPS as measured by compliance with the Compliance Measures (CM) and the program quality scores (CLASS scores); 3

4) The fourth and final section will provide the analyses that produced the key indicators (KI) and recommendations in how it could be used. 2

The source of data for this report is all the Tri-Annual On-Site Monitoring visits for 2012 which consisted of 422 reviews of programs across the country. There were 191 Head Start (HS) only programs, 33 Early Head Start (EHS) only programs, and 198 Head Start/Early Head Start (HS/EHS) programs reviewed. This is a representative sample of Head Start and Early Head Start programs nationally representing approximately 25% of the total number of Head Start programs.

Before proceeding with the results of this study, a few clarifying and definitional terms need to be highlighted. In the 2012 edition of OHS On-Site Review Protocol and the 2013 OHS Monitoring Protocol, Compliance Indicators (CI) and Key Indicators (KI) are respectively mentioned. In the licensing literature, when the term “Indicators” is used it refers to standards/rules that are predictive of overall compliance with all rules/standards. However, as defined by OHS, indicators (CI/KI) are used within the context of risk assessment which means that these indicators are the standards which are most important/critical
to the OHS in their monitoring reviews. These indicators therefore are not predictive in essence. That is the focus of this report/study which is to determine which of these indicators are predictive of overall compliance with all the compliance/key indicators. This is a common misconception in the human service regulatory field where risk assessment tools and key indicator tools purposes are confused. As we move forward please keep the definitions in mind related to the distinctions and functionality of risk assessment and key indicators.

For the purposes of this study, 131 Compliance Measures (CM), organized into seven (7) Content Areas (CA), were reviewed and analyzed. The seven content areas are the following: Program Governance; Management Systems; Fiscal Integrity; Eligibility, Recruitment, Selection, Enrollment, and Attendance; Child Health and Safety; Family and Community Engagement; Child Development and Education. Ten CM’s were from Program Governance (GOV), 10 were from Management Systems (SYS), 22 were from Fiscal Integrity (FIS), 11 were from Eligibility, Recruitment, Selection, Enrollment, and Attendance (ERSEA), 34 were from Child Health and Safety (CHS), 16 were from Family and Community Engagement (FCE), and 28 were from Child Development and Education (CDE).

Section 1 - Head Start (HS), Early Head Start (EHS), and Head Start/Early Head Start (HS/EHS) programs

In order to determine if analyses needed to be performed separately on Head Start (HS), Early Head Start (EHS), and Head Start/Early Head Start (HS/EHS) combined programs, the first series of analyses were performed to determine if any statistically significant differences existed amongst these three groups. This is a very important first analysis because it will help to determine the stability of the sample selected and of the overall system. In other words, is there a good deal of consistency across all service types: HS, EHS, and HS/EHS.

Based upon Table 1, no statistically significant differences were determined amongst the three groups (HS, EHS, HS/EHS) with Compliance Measures (CM) or CLASS (ES, CO, IS) Scores indicating that using the full 422 sample and not having to do separate analyses for the three groups was the correct analytical framework. However, where it is appropriate, any statistically significant differences amongst the various program types will be highlighted.

Table 1 – Head Start, Early Head Start, & Head Start/Early Head Start With CM and CLASS/ES, CO, IS

<table>
<thead>
<tr>
<th>Program Type</th>
<th>CM(N)</th>
<th>CLASS/ES(N)</th>
<th>CLASS/CO(N)</th>
<th>CLASS/IS(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start (HS)</td>
<td>3.72(191)</td>
<td>5.88(186)</td>
<td>5.43(186)</td>
<td>2.97(186)</td>
</tr>
<tr>
<td>Early Head Start (EHS)</td>
<td>2.67(33)</td>
<td>-----*</td>
<td>-----*</td>
<td>-----*</td>
</tr>
<tr>
<td>Head Start (HS/EHS)</td>
<td>3.07(198)</td>
<td>5.91(198)</td>
<td>5.47(198)</td>
<td>3.00(198)</td>
</tr>
<tr>
<td>Totals</td>
<td>3.33(422)</td>
<td>5.89(384)</td>
<td>5.45(384)</td>
<td>2.98(384)</td>
</tr>
</tbody>
</table>

Statistical Significance NS NS NS NS

CM = Compliance Measures (Average Number of Violations)
CLASS/ES = CLASS Emotional Support Average Score
CLASS/CO = CLASS Classroom Organization Average Score
CLASS/IS = CLASS Instructional Support Average Score
NS = Not Significant
N = Number of Programs

*CLASS data were not collected in EHS.
The average number of violations with the Compliance Measures for Head Start (3.72), Early Head Start (2.67) and Head Start/EHS (3.07) was not significant in utilizing a One-Way ANOVA. There were 191 Head Start (HS) programs, 33 Early Head Start (EHS) programs, and 198 Head Start (HS/EHS) programs.

Comparisons were also made with Head Start and Head Start/EHS on the various CLASS sub-scales (ES = Emotional Support, CO = Classroom Organization, and IS = Instructional Support) and no significant differences were found between these two groups. The EHS (n = 33) was not used because CLASS data were not collected in these programs.

The practical implication of the above results is that the same monitoring tools and the resulting Head Start Key Indicator (HSKI) to be developed as a result of this study can be used in the three main types of programs: Head Start, Early Head Start, and Head Start/EHS. There is no need to have separate tools.

**Section 2 - Content Areas**

The second series of analyses was to look more closely at the 7 content areas (CA) to measure demographically any differences amongst the various areas. In order to do this a weighted average had to be determined in order to compare the various areas because of the differences in the number of Compliance Measures (CM) used in each content area. Table 2 provides the results of these analyses.

For the total sample of 422 sites, Management Systems (SYS) Content Area (CA) had the highest number of violations with the Compliance Measures (CM) with 359. The SYS/CA also had the highest average number of violations with 35.90 because there were only 10 CM. For the total sample of 422 sites, the lowest number of violations was in the Family and Community Engagement (FCE) Content Area (CA) with 48 violations with CM. It also had the lowest average number of violations with 3.00.

For the Head Start only sites (n = 191), a similar distribution as with the total sample (n = 422) is depicted in which Management Systems (SYS) Content Area (CA) had the highest number of violations with the Compliance Measures (CM) with 192. The SYS/CA also had the highest average number of violations with 19.20 because again there were only 10 CM. The lowest number of violations was in the Family and Community Engagement (FCE) Content Area (CA) with 20 violations with CM. It also had the lowest average number of violations with 1.25.

For the Early Head Start only (n = 33) and the Head Start/Early Head Start (n = 198) sites, the ranking of the various Content Areas changed somewhat with the total number of violations and the average number of violations from the Total Sample (n = 422) and the Head Start only (n = 191) sites but not dramatically. For example, the Family and Community Engagement (FCE); Child Development and Education (CDE); and the Eligibility, Recruitment, Selection, Enrollment, and Attendance (ERSEA) Content Areas switched rankings in which it had the fewest total violations and the average number of violations (see Table 2).
Table 2 – Comparing Content Areas and Program Types

<table>
<thead>
<tr>
<th>Content Areas</th>
<th>Total Violations/(Rank)</th>
<th>Average # of Violations/(Rank)</th>
<th>CM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOT</td>
<td>HS</td>
<td>EHS</td>
</tr>
<tr>
<td>FCE</td>
<td>48(1)</td>
<td>20(1)</td>
<td>2(1)</td>
</tr>
<tr>
<td>ERSEA</td>
<td>62(2)</td>
<td>37(2)</td>
<td>6(3)</td>
</tr>
<tr>
<td>CDE</td>
<td>91(3)</td>
<td>43(3)</td>
<td>5(2)</td>
</tr>
<tr>
<td>GOV</td>
<td>150(4)</td>
<td>94(4)</td>
<td>6(3)</td>
</tr>
<tr>
<td>FIS</td>
<td>255(5)</td>
<td>114(5)</td>
<td>23(7)</td>
</tr>
<tr>
<td>CHS</td>
<td>333(6)</td>
<td>151(6)</td>
<td>22(6)</td>
</tr>
<tr>
<td>SYS</td>
<td>359(7)</td>
<td>192(7)</td>
<td>20(5)</td>
</tr>
</tbody>
</table>

**CONTENT AREAS (CA):**
FCE = FAMILY and COMMUNITY ENGAGEMENT
ERSEA = ELIGIBILITY, RECRUITMENT, SELECTION, ENROLLMENT, and ATTENDANCE
CDE = CHILD DEVELOPMENT AND EDUCATION
GOV = PROGRAM GOVERNANCE
FIS = FISCAL INTEGRITY
CHS = CHILD HEALTH AND SAFETY
SYS = MANAGEMENT SYSTEMS

**TOT** = TOTAL NUMBER OF SITES, FULL SAMPLE OF 422 SITES
**HS** = HEAD START ONLY PROGRAMS
**EHS** = EARLY HEAD START ONLY PROGRAM
**HS/EHS** = HEAD START AND EARLY HEAD START COMBINED PROGRAMS
**CM** = NUMBER OF COMPLIANCE MEASURES

**TOTAL VIOLATIONS** = ALL THE VIOLATIONS FOR A SPECIFIC CONTENT AREA.
**AVERAGE # OF VIOLATIONS** = THE TOTAL VIOLATIONS FOR A SPECIFIC CA DIVIDED BY THE NUMBER OF COMPLIANCE MEASURES FOR THAT SPECIFIC CONTENT AREA.
**RANK** = HOW EACH CONTENT AREA COMPARES TO THE OTHER CONTENT AREAS FOR THE RESPECTIVE PROGRAM TYPE.

For the total sample (n = 422), other CA’s had different configurations between the total number of violations and the average number of violations as demonstrated by CHS – Child Health and Safety in which there was a total of 333 violations but the average number of violations was 9.79 because there were 34 Compliance Measures (CM). Program Governance (GOV) had 150 total violations and a weighted-average of 15 violations with 10 CM. Child Development and Education (CDE) had 91 total violations and a weighted-average of 3.25 violations. Fiscal Integrity (FIS) had 255 total violations and a weighted-average of 11.59 violations. And lastly, Eligibility, Recruitment, Selection, Enrollment, and Attendance (ERSEA) had 62 total violations and a weighted-average of 5.64 violations.

The Head Start only (HS = 191), Early Head Start only (EHS = 33), and the Head Start/Early Head Start (HS/EHS = 198) programs followed a similar pattern as with the total sample (n = 422). This indicates a great deal of consistency in the sample drawn. See Appendix 4 for violation data for all 131 Compliance Measures.

The practical implication of the above findings is that certain Content Areas (SYS, GOV, FIS) may need additional exploration by OHS because of their high rates of non-compliance with the Compliance Measures.
Section 3 – Program Quality

This section provides comparisons between the Compliance Measures (CM) data and the CLASS (ES, CO, IS) data. This is a very important section because there is always the concern that compliance with the HSPS has no relationship to program quality as measured by the CLASS. In Table 3, correlations were run between the CM data and the CLASS scores for Emotional Support (ES), Classroom Organization (CO), and Instruction Support (IS) for the Head Start only and the Head Start/Early Head Start programs. The EHS only programs were not included because CLASS data are not collected on these programs. The results are very positive and statistically significant in most cases. It is also important to note the very positive correlation between the Head Start Key Indicators (HSKI) and CLASS. This result supports using the HSKI in monitoring Head Start.

Table 3 – Relationship Between Compliance Measures (CM), KI, and CLASS (ES, CO, IS) Scores

<table>
<thead>
<tr>
<th>Compliance Measures Content Areas</th>
<th>Key Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS/ES</td>
<td>CM</td>
</tr>
<tr>
<td>CLASS/CO</td>
<td>.19**</td>
</tr>
<tr>
<td>CLASS/IS</td>
<td>.20**</td>
</tr>
</tbody>
</table>

CM Violations = Total Compliance Measure Violations

** p < .01
* p < .05

CONTENT AREAS (CA):
FCE = FAMILY and COMMUNITY ENGAGEMENT
ERSEA = ELIGIBILITY, RECRUITMENT, SELECTION, ENROLLMENT, and ATTENDANCE
CDE = CHILD DEVELOPMENT AND EDUCATION
GOV = PROGRAM GOVERNANCE
FIS = FISCAL INTEGRITY
CHS = CHILD HEALTH AND SAFETY
SYS = MANAGEMENT SYSTEMS

CLASS/IS = Average CLASS IS (Instructional Support) Score
CLASS/ES = Average CLASS ES (Emotional Support) Score
CLASS/CO = Average CLASS CO (Classroom Organization) Score

KI = Key Indicators Total Score

See Appendix 6 & 6A for the inter-correlations amongst all the Content Areas, HSKI, and Total Compliance with Compliance Measures.

These results are very important but it is equally important to look more specifically at the distribution of the Compliance Measures (CM) scores and their relationship to the CLASS data (see Appendix 5 for detailed graphic distributions and Appendix 6 & 6A for the inter-correlations amongst all the CA). When this is done a very interesting trend appears (see Table 3a) in which a definite plateau occurs as the scores move from more violations or lower compliance with the Compliance Measures (25-20 to 3-8 CM Violations) to fewer violations or substantial compliance with the Compliance Measures (1-2 CM Violations) and full compliance with the Compliance Measures (Zero (0) CM Violations).
When comparing these groupings in Table 3a the results from a One Way ANOVA were significant (F = 4.92; p < .001) for the CLASS/IS Scores. The average CLASS/IS Score when there were no CM Violations was 3.03. The average CLASS/IS Score when there were 1-2 CM Violations was 3.15. The average CLASS/IS Score when there were 3-8 CM Violations was 2.87. The average CLASS/IS Score when there were 9-19 CM Violations was 2.65. And finally, the average CLASS/IS Score when there were 20-25 violations was 2.56. The results were very similar with the CLASS/ES and CLASS/CO scores as well in which the results from a One Way ANOVA were statistically significant for the CLASS/ES (F = 4.918; p < .001) and for the CLASS/CO (F = 4.174; p < .003). These results clearly demonstrate that being in full or substantial compliance with the Compliance Measures correlates with more positive scores on the CLASS. Approximately 55% of the Head Start programs are at the full or substantial compliance level.

The practical implication of the above findings is that placing equal emphasis on full as well as substantial compliance with the Compliance Measures could be an acceptable public policy decision.

### Section 4 – Head Start Key Indicators (HSKI)

The fourth and final section of this report is in some ways the most important since this is the focus of the study: developing statistically predictive Key Indicator (KI) Compliance Measures (CM) – the Head Start Key Indicators (HSKI).

These are the statistically predictive Key Indicators based upon the KI methodology, correlations with the CLASS/ES, CO, IS, and correlations with the CM Total Violation scores. Table 4 lists the results while Appendix 1 has the specific KI’s content specified. Appendix 2 depicts the KI Formula Matrix. Only those Compliance Measures (CM) that had significant results on three of the five correlations were selected to be Head Start Key Indicator Compliance Measures (HSKI).

The methodology used to generate the Compliance Measure Key Indicators sorted the top 20% of programs in compliance and compared this group to the bottom 27% of programs in compliance. The middle 53% of programs were not used in order to determine the Key Indicators. These cut off points

---

**Table 3a – Aggregate Scores Comparing CM Violations with CLASS Scores**

<table>
<thead>
<tr>
<th>CM Violations</th>
<th>IS</th>
<th>ES</th>
<th>CO</th>
<th>Number/Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Full Compliance)</td>
<td>3.03</td>
<td>5.99</td>
<td>5.59</td>
<td>75/19%</td>
</tr>
<tr>
<td>1-2 (Substantial Compliance)</td>
<td>3.15</td>
<td>5.93</td>
<td>5.50</td>
<td>135/35%</td>
</tr>
<tr>
<td>3-8 (Mid-Compliance)</td>
<td>2.87</td>
<td>5.85</td>
<td>5.37</td>
<td>143/40%</td>
</tr>
<tr>
<td>9-19 (Lower Compliance)</td>
<td>2.65</td>
<td>5.71</td>
<td>5.32</td>
<td>28/6%</td>
</tr>
<tr>
<td>20-25 (Lowest Compliance)</td>
<td>2.56</td>
<td>5.52</td>
<td>4.93</td>
<td>3/1%</td>
</tr>
</tbody>
</table>

Significance: IS = F = 4.92; p < .001, ES = F = 4.918; p < .001, CO = F = 4.174; p < .003

**CM Violations = Compliance Measure Violations (lower score = higher compliance)(higher score = lower compliance)
IS = Average CLASS IS (Instructional Support) Score
ES = Average CLASS ES (Emotional Support) Score
CO = Average CLASS CO (Classroom Organization) Score
#/% = Number of programs and Percent of programs at each level of compliance**
were determined by the compliance distribution in which 20% of the programs were in 100% compliance while 27% of the programs had compliance scores of 95% or less.

Table 4 – Head Start Key Indicator (HSKI) Compliance Measures (CM) and CLASS and Total Violations

<table>
<thead>
<tr>
<th>HSKI/CM (2013)</th>
<th>Phi</th>
<th>CLASS/ES</th>
<th>CLASS/CO</th>
<th>CLASS/IS</th>
<th>Total Violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDE4.1</td>
<td>.28***</td>
<td>.10*</td>
<td>ns</td>
<td>ns</td>
<td>.30***</td>
</tr>
<tr>
<td>CHS1.1</td>
<td>.39***</td>
<td>.15**</td>
<td>.16**</td>
<td>.10*</td>
<td>.39***</td>
</tr>
<tr>
<td>CHS1.2</td>
<td>.33***</td>
<td>.18**</td>
<td>.15**</td>
<td>ns</td>
<td>.36***</td>
</tr>
<tr>
<td>CHS2.1</td>
<td>.49***</td>
<td>.18**</td>
<td>.15**</td>
<td>ns</td>
<td>.54***</td>
</tr>
<tr>
<td>CHS3.10</td>
<td>.39***</td>
<td>.11*</td>
<td>.11*</td>
<td>ns</td>
<td>.24***</td>
</tr>
<tr>
<td>GOV2.1</td>
<td>.31***</td>
<td>.11*</td>
<td>ns</td>
<td>ns</td>
<td>.46***</td>
</tr>
<tr>
<td>SYS2.1</td>
<td>.47***</td>
<td>.15**</td>
<td>.16**</td>
<td>.14**</td>
<td>.55***</td>
</tr>
<tr>
<td>SYS3.4</td>
<td>.58***</td>
<td>.13*</td>
<td>.10*</td>
<td>ns</td>
<td>.36***</td>
</tr>
</tbody>
</table>

* Phi = the phi coefficient which statistically predicts compliance with the full set of CM’s.
** CLASS/ES = correlations between the specific CM and this specific scale of the CLASS.
*** CLASS/CO = correlations between the specific CM and this specific scale of the CLASS.
### Total Violations = correlations between the specific CM and the total number of CM violations for each program.

Separate Key Indicators were run for just Head Start only and Head Start/Early Head Start programs but the key indicators were only a subset of the above list, albeit a shorter list in each case. Based upon those phi coefficients, it was determined that using the above list for all Head Start only, Early Head Start, and Head Start/Early Head Start was a more efficient and effective way to monitor all the programs with one list of indicators rather than having separate key indicators for program types. The separate phi coefficients run for Head Start only and Head Start/Early Head Start programs did not show any significant differences because they were sub-samples of the overall sample drawn.

Section 4A – Suggested Use of the HSKI for Head Start Program Monitoring

Now that Key Indicators have been generated, the next question is how to use HSKI in the program monitoring of Head Start. A possible way in which the HSKI could be used would be the following (see Figure 1) in which a differential monitoring approach could be used:

All programs would be administered the HSKI. If there is full (100%) compliance with the Head Start Key Indicators (HSKI) then the next scheduled review of the program would be an Abbreviated Monitoring Visit (AMV). If there is not 100% compliance with the Head Start Key Indicators (HSKI) then the next scheduled review of the program would be a Full Monitoring Visit (FMV) in which all Compliance Measures are reviewed. Based upon the results of the FMV a determination could be made regarding a compliance or non-compliance decision (see Figure 1) and how often the program will be visited.
Figure 1 – Head Start Key Indicator (HSKI) Compliance Measures Differential Monitoring Model

Compliance Decisions:

Head Start Key Indicators (HSKI) – this becomes a screening tool to determine if a program receives an AMV OR FMV visit.

HSKI (100%) = For the next visit, an Abbreviated Monitoring Visit (AMV) is conducted. Every 3-4 yrs a full Monitoring is conducted.

HSKI (not 100%) = For the next visit, a Full Monitoring Visit (FMV) is conducted and all CMs are reviewed.

Compliance = 98%+ with all CMs which indicates substantial to full compliance and 100% with HSKI. For the next visit, an Abbreviated Monitoring Visit (AMV) is conducted.

Non-compliance = less than 98% with all CMs which indicates low compliance. For the next visit a Full Monitoring Visit (FMV) is conducted.

Moving to a differential monitoring system could provide a cost effective and efficient model for Head Start program monitoring. This revision to the Head Start program monitoring system would combine a risk assessment and key indicator approach (see Appendix 3) in determining what compliance measures to review, how often, and how comprehensive a review should be utilized. It would continue to focus on the most critical compliance measures that statistically predict overall compliance with the full complement of compliance measures.

See Appendix 7 – Figure 2 for how the above differential monitoring system could impact the present Head Start Tri-Annual Review Monitoring System. In this appendix, a cost neutral monitoring system is proposed based upon the above DMLMA/Key Indicator Model.
References


Footnotes

1) PIR Dashboard Key Indicators could not be generated because the PIR data demonstrated little statistical predictive ability to be useful for discriminating between high and low compliant programs or program quality with the exception of staff having CDA’s.

2) The correlation between Compliance Measures (CM) and the statistically predictive Key Indicators (HSKI) was .77 which exceeds the expected correlation threshold.

3) The correlations between the CLASS/ES, CO, IS and Key Indicators were the following: .27, .25, .17 respectively. The correlations between KI and ES, CO were higher than the correlations between CM and ES, CO as reported earlier in this report. The correlation between IS and CM was higher .20 than KI and IS (.17).

4) Because this study spans the 2012 Review Protocol and 2013 Monitoring Protocol, Compliance Indicators and Compliance Measures are used interchangeably with a preference given to using Compliance Measures (CM) in this report. There are 139 Compliance Indicators; 115 Compliance Measures, but for the purposes of this study 131 Compliance Measures were available in the 2012 Head Start data base drawn for this study.

For additional information regarding this report, please contact:
Richard Fiene, Ph.D., Director
Research Institute for Key Indicators
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
RIKI.Institute@gmail.com
717-944-5868 Phone and Fax
http://RIKInstitute.wikispaces.com
http://pennstate.academia.edu/RickFiene

February 2013 (revised March, April 2013)
# Appendix 1 – Head Start Key Indicators (HSKI) Compliance Measures Content

<table>
<thead>
<tr>
<th>CM</th>
<th>Content</th>
<th>Regulations/Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHS1.1</td>
<td>The program engages parents in obtaining from a health care professional a determination of whether each child is up to date on a schedule of primary and preventive health care (including dental) and assists parents in bringing their children up to date when necessary and keeping their children up to date as required.</td>
<td>1304.20(a)(1)(ii), 1304.20(a)(1)(ii)(A), 1304.20(a)(1)(ii)(B)</td>
</tr>
<tr>
<td>CHS1.2</td>
<td>The program ensures that each child with a known, observable, or suspected health, oral health, or developmental problem receives follow-up and further testing, examination, and treatment from a licensed or certified health care professional.</td>
<td>1304.20(a)(1)(iii), 1304.20(a)(1)(iv), 1304.20(c)(3)(ii)</td>
</tr>
<tr>
<td>CHS2.1</td>
<td>The program, in collaboration with each child’s parent, performs or obtains the required linguistically and age-appropriate screenings to identify concerns regarding children within 45 calendar days of entry into the program, obtains guidance on how to use the screening results, and uses multiple sources of information to make appropriate referrals.</td>
<td>1304.20(a)(2), 1304.20(b)(1), 1304.20(b)(2), 1304.20(b)(3)</td>
</tr>
<tr>
<td>CHS3.10</td>
<td>Maintenance, repair, safety of facility and equipment</td>
<td>1304.53(a)(7)</td>
</tr>
<tr>
<td>GOV2.1*</td>
<td>Members of the governing body and the Policy Council receive appropriate training and technical assistance to ensure that members understand information they receive and can provide effective oversight of, make appropriate decisions for, and participate in programs of the Head Start agency.</td>
<td>642(d)(3)</td>
</tr>
<tr>
<td>SYS2.1</td>
<td>The program established and regularly implements a process of ongoing monitoring of its operations and services, including delegate agencies, in order to ensure compliance with Federal regulations, adherence to its own program procedures, and progress towards the goals developed through its Self-Assessment process.</td>
<td>1304.51(i)(2), 641A(g)(3)</td>
</tr>
<tr>
<td>SYS3.4</td>
<td>Prior to employing an individual, the program obtains a: Federal, State, or Tribal criminal record check covering all jurisdictions where the program provides Head Start services to children; Federal, State, or Tribal criminal record check as required by the law of the jurisdiction where the program provides Head Start services; Criminal record check as otherwise required by Federal law</td>
<td>648A(g)(3)(A), 648A(g)(3)(B), 648A(g)(3)(C)</td>
</tr>
</tbody>
</table>

Appendix 2: Key Indicator Formula Matrix for HSKI – Head Start Key Indicators

<table>
<thead>
<tr>
<th></th>
<th>Providers In Compliance</th>
<th>Programs Out Of Compliance</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Group</strong></td>
<td><strong>A</strong></td>
<td><strong>B</strong></td>
<td><strong>Y</strong></td>
</tr>
<tr>
<td><strong>Low Group</strong></td>
<td><strong>C</strong></td>
<td><strong>D</strong></td>
<td><strong>Z</strong></td>
</tr>
<tr>
<td><strong>Column Total</strong></td>
<td><strong>W</strong></td>
<td><strong>X</strong></td>
<td><strong>Grand Total</strong></td>
</tr>
</tbody>
</table>

Key Indicator Statistical Methodology (Calculating the Phi Coefficient):

\[
\phi = \frac{(A)(D) - (B)(C)}{\sqrt{(W)(X)(Y)(Z)}}
\]

A = High Group + Programs in Compliance on Specific Compliance Measure.
B = High Group + Programs out of Compliance on Specific Compliance Measure.
C = Low Group + Programs in Compliance on Specific Compliance Measure.
D = Low Group + Programs out of Compliance on Specific Compliance Measure.

W = Total Number of Programs in Compliance on Specific Compliance Measure.
X = Total Number of Programs out of Compliance on Specific Compliance Measure.
Y = Total Number of Programs in High Group.
Z = Total Number of Programs in Low Group.

High Group = Top 20% of Programs in Compliance with all Compliance Measures.
Low Group = Bottom 27% of Programs in Compliance with all Compliance Measures.

<table>
<thead>
<tr>
<th>Phi Coefficient Range</th>
<th>Characteristic of Indicator</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+1.00) – (+.26)</td>
<td>Good Predictor</td>
<td>Include on HSKI</td>
</tr>
<tr>
<td>(+.25) – (0)</td>
<td>Too Easy</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(0) – (-.25)</td>
<td>Too Difficult</td>
<td>Do not Include</td>
</tr>
<tr>
<td>(-.26) – (-1.00)</td>
<td>Terrible Predictor</td>
<td>Do not Include</td>
</tr>
</tbody>
</table>
Appendix 3

DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (Fiene, 2012) DMLMA© Applied to the Office of Head Start Program Monitoring Compliance System

CI + PQ => RA + KI => DM

*Head Start Examples:*

CI = Head Start Performance Standards (HSPS)
PQ = CLASS ES, IS, CO (CLASS)
RA = Compliance Measures (CM)
KI = Key Indicators (generated from this study = Head Start Key Indicators (HSKI))
DM = Not Applicable at this time (NA) but see Figure 1 for a proposed model

*DMLMA© Thresholds:*

*High Correlations (.70+)* = CI x KI.
*Moderate Correlations (.50+) =* CI x RA; RA x DM; RA x KI; KI x DM.
*Lower Correlations (.30+) =* PQ x CI; PQ x RA; PQ x KI.
## Appendix 4: Content Areas and Compliance Measures

<table>
<thead>
<tr>
<th>Content Areas and Compliance Measures</th>
<th>Percent (%) Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CDE - CHILD DEVELOPMENT AND EDUCATION</strong></td>
<td></td>
</tr>
<tr>
<td>1.1.(2.2) The program implements a curriculum that is aligned with the Head Start Child Development and Early Learning Framework...</td>
<td>99%</td>
</tr>
<tr>
<td>1.2 The program implements a curriculum that is evidence-based...</td>
<td>99%</td>
</tr>
<tr>
<td>1.3.(2.1) The curriculum is comprehensive...</td>
<td>99%</td>
</tr>
<tr>
<td>2.1 The program implements an infant toddler curriculum...</td>
<td>99%</td>
</tr>
<tr>
<td>2.2 The program develops secure relationships in out of home care settings for infants and toddlers...</td>
<td>100%</td>
</tr>
<tr>
<td>2.3 The program implements an infant/toddler curriculum that encourages trust...</td>
<td>100%</td>
</tr>
<tr>
<td>2.4 The program encourages the development of self-awareness, autonomy...</td>
<td>100%</td>
</tr>
<tr>
<td>2.5 The program fosters independence.</td>
<td>100%</td>
</tr>
<tr>
<td>2.6 The program enhances each child’s strengths by encouraging self control...</td>
<td>99%</td>
</tr>
<tr>
<td>2.7 The program plans for routines and transitions...</td>
<td>99%</td>
</tr>
<tr>
<td>2.8 The program ensures respect for others feelings and rights.</td>
<td>99%</td>
</tr>
<tr>
<td>2.10 The program provides opportunities for children to engage in child-initiated...</td>
<td>100%</td>
</tr>
<tr>
<td>2.11 Nutrition services contribute to children’s development and socialization...</td>
<td>100%</td>
</tr>
<tr>
<td>3.1 The program uses information from screenings, ongoing observations...</td>
<td>99%</td>
</tr>
<tr>
<td>3.2 The programs’ nutrition program is designed and implemented to meet the nutritional needs...</td>
<td>98%</td>
</tr>
<tr>
<td>3.4(CHS4.5) Meal and snack periods are appropriately scheduled...</td>
<td>99%</td>
</tr>
<tr>
<td>3.5.(3.2) Services provided to children with identified disabilities are designed to support...</td>
<td>100%</td>
</tr>
<tr>
<td>3.6(3.3) The program designates a staff member or consultant to coordinate services for children w/disabilities...</td>
<td>100%</td>
</tr>
<tr>
<td>3.7(3.4) The program has secured the services of a mental health professional...</td>
<td>97%</td>
</tr>
<tr>
<td>3.8(3.5) The program’s approach to CDE is developmentally and linguistically appropriate...</td>
<td>99%</td>
</tr>
<tr>
<td>4.1 The program establishes goals for improving school readiness...</td>
<td>98%</td>
</tr>
<tr>
<td>4.2 The program uses self assessment information on school readiness goals...</td>
<td>99%</td>
</tr>
<tr>
<td>4.3 The program demonstrates that children who are dual language learners...</td>
<td>100%</td>
</tr>
<tr>
<td>5.1(4.1) The program hires teachers who have the required qualifications, training, &amp; experience.</td>
<td>92%</td>
</tr>
<tr>
<td>5.2 The program ensures that family child care providers have the required qualifications...</td>
<td>100%</td>
</tr>
<tr>
<td>5.3 The program ensures that all full time Head Start employees who provide direct education...</td>
<td>96%</td>
</tr>
<tr>
<td>5.4 The program ensures that home visitors have the required qualifications, training...</td>
<td>99%</td>
</tr>
<tr>
<td>5.5 When the majority of children speak the same language...</td>
<td>99%</td>
</tr>
<tr>
<td><strong>CHS - CHILD HEALTH AND SAFETY</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 The program engages parents in obtaining from a health care professional a determination of whether each child...</td>
<td>89%</td>
</tr>
<tr>
<td>1.2 The program ensures that each child with a known, observable, or suspected health, oral health...</td>
<td>92%</td>
</tr>
<tr>
<td>1.3 The program involves parents, consulting with them immediately when child health or developmental problems...</td>
<td>100%</td>
</tr>
<tr>
<td>1.4 The program informs parents and obtains authorization prior to all health procedures...</td>
<td>98%</td>
</tr>
<tr>
<td>1.5 The program has established procedures for tracking the provision of health services.</td>
<td>97%</td>
</tr>
<tr>
<td>1.6 The EHS program helps pregnant women, immediately after enrollment in the program, access through referrals...</td>
<td>100%</td>
</tr>
<tr>
<td>1.7 Program health staff conduct a home visit or ensure that a health staff member visits each newborn within 2 weeks of birth...</td>
<td>97%</td>
</tr>
<tr>
<td>2.1 The program, in collaboration with each child’s parent, performs or obtains the required screenings...</td>
<td>84%</td>
</tr>
<tr>
<td>2.2 A coordinated screening, assessment, and referral process for all children...</td>
<td>98%</td>
</tr>
<tr>
<td>2.3 The program, in partnership with the LEA or Part C Agency, works to inform and engage parents in all plans for screenings...</td>
<td>99%</td>
</tr>
<tr>
<td>3.1 Facilities used for center based program options comply with state and local licensing...</td>
<td>100%</td>
</tr>
<tr>
<td>3.2 The program ensures that sufficient equipment, toys, materials, and furniture are provided...</td>
<td>97%</td>
</tr>
<tr>
<td>3.3 Precautions are taken to ensure the safety of children.</td>
<td>99%</td>
</tr>
<tr>
<td>3.4 The program ensures that medication is properly stored and is not accessible to children.</td>
<td>98%</td>
</tr>
<tr>
<td>3.5 The program ensures that no hazards are present around children.</td>
<td>89%</td>
</tr>
<tr>
<td>3.6 The program ensures that sleeping arrangements for infants do not use soft bedding materials.</td>
<td>99%</td>
</tr>
<tr>
<td>3.7 All infant and toddler toys are made of non-toxic materials and sanitized regularly.</td>
<td>99%</td>
</tr>
<tr>
<td>3.8 The program has adequate usable indoor and outdoor space.</td>
<td>99%</td>
</tr>
<tr>
<td>3.9 Outdoor play areas are arranged to prevent children from getting into unsafe or unsupervised areas...</td>
<td>100%</td>
</tr>
<tr>
<td>3.10 The program provides for maintenance, repair, safety, and security of all Head Start facilities and equipment.</td>
<td>85%</td>
</tr>
<tr>
<td>3.11 The program’s facilities provide adequately for children with disabilities...</td>
<td>100%</td>
</tr>
<tr>
<td>4.1 Staff, volunteers, and children wash their hands with soap and running water.</td>
<td>98%</td>
</tr>
<tr>
<td>4.2 Spilled bodily fluids are cleaned up and disinfected immediately.</td>
<td>100%</td>
</tr>
<tr>
<td>4.3 The program adopts sanitation and hygiene practices for diapering...</td>
<td>99%</td>
</tr>
</tbody>
</table>
4.4(4.7) The program ensures that facilities are available for proper refrigerated storage and handling of breast milk and formula.... 100%
4.5(4.8) Effective oral hygiene is promoted among children in conjunction with meals.... 99%
5.1 The program ensures appropriate class and group sizes based on the predominant age of the children.... 99%
5.2 The program ensures that no more than eight children are placed in an infant and toddler space.... 99%
6.1 The program’s vehicles are properly equipped.... 99%
6.2 At least one bus monitor is aboard the vehicle at all times.... 99%
6.3 Children are released only to a parent.... 99%
6.4 Each bus monitor, before duty, has been trained on child boarding and exiting procedures.... 99%
6.5 The program ensures that persons employed to drive vehicles receive the required behind the wheel training.... 99%
6.6 Specific types of transportation assistance offered are made clear to all prospective families.... 100%

ERSEA – ELIGIBILITY, RECRUITMENT, SELECTION, ENROLLMENT, AND ATTENDANCE 98%
1.1 The program developed and implemented a process that is designed to actively recruit families.... 99%
1.2 The program has a systematic process for establishing selection criteria.... 99%
1.3 The program has established and implemented outreach and enrollment policies and procedures.... 99%
2.1 Program staff verified each child’s eligibility.... 94%
2.2 The program enrolls children who are categorically eligible.... 99%
2.3 The American Indian or Alaskan Native programs ensure that the children who meet the following requirements.... 100%
3.1 Actual program enrollment is composed of at least 10 percent children with disabilities.... 96%
3.2 The program enrolled 100% of its funded enrollment.... 98%
3.3 The program has documentation to support monthly enrollment data.... 98%
4.1 When monthly average daily attendance in center based programs falls below 85%, the causes of absenteeism.... 99%
4.2 The program ensures that no child’s enrollment or participation in the Head Start program is contingent on payment of a fee.... 99%

FCE – FAMILY AND COMMUNITY ENGAGEMENT 99%
1.1(1.2) Program staff are familiar with the backgrounds of families and children.... 100%
1.2(1.3) A strength based and family driven collaborative partnership building process is in place.... 100%
1.3(1.4) The program provides resources and services for families’ needs, goals, and interests.... 99%
2.1 The program provides opportunities for parents to enhance their parenting skills.... 99%
2.2 Parents and staff share their respective concerns and observations about their individual children.... 99%
2.3 On site mental health consultation assists the program in providing education to parents.... 97%
3.1 Program staff plan, schedule, and facilitate no fewer than two staff parent conferences.... 98%
3.2(1.1) The program is open to parents during all program hours.... 99%
3.3(1.2) In home based settings, programs encourage parents to be integrally involved in their children’s development.... 99%
3.4(1.3) Programs provide opportunities for children and families to participate in literacy services.... 99%
3.5(1.4) The program builds parents’ confidence to advocate for their children by informing parents of their rights.... 99%
4.1 The program has procedures to support successful transitions for enrolled children.... 99%
4.2 The program initiates transition planning for each EHS enrolled child at least 6 months prior to the child’s 3rd birthday.... 99%
5.1 The program has established and maintains a health services advisory committee.... 97%
5.2 The program has taken steps to establish ongoing collaborative relationships with community organizations.... 100%
5.3 The program coordinates with and has current interagency agreements in place with LEA’s.... 98%

FIS – FISCAL INTEGRITY 97%
1.1 The program’s financial management systems provide for effective control.... 94%
1.2 The program sought and received prior approval in writing for budget changes.... 99%
1.3 The program minimized the time elapsing between the advancement of funds from the Payment Management System.... 100%
1.4 The program used Head Start funds to pay the cost of expenses.... 99%
1.5 The program has obtained and maintained required insurance coverage for risks and liabilities.... 99%
2.1 Financial reports and accounting records are current, accurate, complete.... 98%
2.2 Monthly financial statements, are provided to program governing bodies and policy groups.... 97%
3.1(1.3) The program has procurement procedures that provide all requirements specified in the applicable statutes.... 95%
3.2(1.4) Contracts and delegate agency agreements are current, available, signed, and dated.... 96%
4.1 Original time records are prepared and properly signed by the individual employee & approved.... 97%
4.2 Head Start or EHS grant funds are not used as any part of the monetary compensation.... 99%
4.3 Total compensation for personal services charged to the grant are allowable and reasonable.... 98%
5.1 The grantee has implemented procedures to determine allowable, allocability, and reasonableness of costs.... 95%
5.2 Indirect cost charges are supported by a negotiated and approved indirect cost rate.... 100%
5.3 If the grantee is required to allocate costs between funding sources, the program utilizes a method for allocating costs.... 97%
5.4 The financial records of the grantee are sufficient to allow verification that non-Federal participation is necessary.... 90%
5.5(5.5) The grantee can demonstrate that all contributions of non-Federal share are necessary and reasonable.... 98%
5.6(5.4) During each funding period reviewed the grantee charged to the award only costs resulting from obligations.... 98%
6.1(6.1) For grantees that own facilities purchased or constructed using Head Start grant funds, documentation is available.... 97%
6.2(6.1) The grantee meets property management standards for equipment purchased using HS funds.... 94%
6.3(6.1) Grantees that entered into a mortgage or other loan agreement using collateral property complied with Federal regs.... 97%
6.4(6.1) The amount which the grantee may claim a cost or non-Federal share contribution.... 96%

GOV – PROGRAM GOVERNANCE 96%
<table>
<thead>
<tr>
<th></th>
<th>The program has a governing body.</th>
<th>98%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>The program has established a policy council.</td>
<td>98%</td>
</tr>
<tr>
<td>2.1</td>
<td>Policy council and policy committee members are supported by the program.</td>
<td>99%</td>
</tr>
<tr>
<td>2.2</td>
<td>The program has policies and procedures in place to ensure that member of the governing body &amp; PAC are free.</td>
<td>97%</td>
</tr>
<tr>
<td>3.1(2.1)</td>
<td>Members of the governing body and the PAC receive appropriate training and TA.</td>
<td>94%</td>
</tr>
<tr>
<td>3.2(2.2)</td>
<td>The governing body performs required activities and makes decisions pertaining to program administration.</td>
<td>95%</td>
</tr>
<tr>
<td>3.3</td>
<td>The governing body approves financial management, accounting, and reporting policies.</td>
<td>99%</td>
</tr>
<tr>
<td>3.4</td>
<td>The governing body reviews and approves all of the program’s major policies.</td>
<td>95%</td>
</tr>
<tr>
<td>3.5(2.4)</td>
<td>The PAC approves and submits decisions about identified program activities to the governing body.</td>
<td>98%</td>
</tr>
<tr>
<td>4.1(3.1)</td>
<td>Governing body and PAC members regularly receive and use information about program planning.</td>
<td>88%</td>
</tr>
</tbody>
</table>

**SYS – MANAGEMENT SYSTEMS**

|   | The program routinely engages in a process of systematic planning that utilizes the results of the community assessment. | 97% |
| 1.2(5.1) | At least annually, the program conducts a self assessment of program effectiveness. | 97% |
| 2.2 | The program established and maintains a record keeping system regarding children, families, and staff. | 92% |
| 2.3 | The program publishes and makes available to the public an annual report. | 88% |
| 3.1 | The program has established an organizational structure that provides for adequate supervision. | 97% |
| 3.2 | The program develops and implements written standards of conduct. | 97% |
| 3.3 | The program ensures that each staff member completes an initial health examination. | 90% |
| 3.4 | Prior to employing an individual, the program obtains: criminal record check. | 66% |
| 4.1 | The program has mechanisms for regular communication among all program staff. | 98% |
Appendix 5 – Histograms of Total Compliance Measure Violations, CLASS (IS, ES, CO) Scores and Head Start Key Indicator (HSKI) Scores

Total Compliance Measure Violations

Mean = 3.33
Std. Dev. = 3.769
N = 422
CLASS ES Scores

Mean = 5.8935
Std. Dev. = .3578
N = 384
CLASS CO Scores

Mean = 5.4506
Std. Dev. = .4905
N = 384
CLASS IS Scores

Mean = 2.9033
Std. Dev. = .7030
N = 384
Head Start Key Indicators (HSKI) Scores

Mean = 1.00
Std. Dev. = 1.365
N = 422
Appendix 6 -

CONTENT AREA (CA)
CORRELATIONS

<table>
<thead>
<tr>
<th></th>
<th>CHS</th>
<th>ERSEA</th>
<th>FCE</th>
<th>FIS</th>
<th>GOV</th>
<th>SYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDE</td>
<td>.33**</td>
<td>.26**</td>
<td>.06</td>
<td>.14**</td>
<td>.13*</td>
<td>.33**</td>
</tr>
<tr>
<td>CHS</td>
<td></td>
<td>.29**</td>
<td>.18**</td>
<td>.09</td>
<td>.25**</td>
<td>.51**</td>
</tr>
<tr>
<td>ERSEA</td>
<td></td>
<td></td>
<td>.15**</td>
<td>.10*</td>
<td>.27**</td>
<td>.38**</td>
</tr>
<tr>
<td>FCE</td>
<td></td>
<td></td>
<td></td>
<td>.25**</td>
<td>.51**</td>
<td></td>
</tr>
<tr>
<td>FIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.38**</td>
<td></td>
</tr>
</tbody>
</table>

* P < .05
** P < .01

CONTENT AREAS (CA):
FCE = FAMILY and COMMUNITY ENGAGEMENT
ERSEA = ELIGIBILITY, RECRUITMENT, SELECTION, ENROLLMENT, and ATTENDANCE
CDE = CHILD DEVELOPMENT AND EDUCATION
GOV = PROGRAM GOVERNANCE
FIS = FISCAL INTEGRITY
CHS = CHILD HEALTH AND SAFETY
SYS = MANAGEMENT SYSTEMS

Appendix 6A – Total Compliance with Compliance Measures, HSKI, and Content Area Correlations

<table>
<thead>
<tr>
<th></th>
<th>TOT</th>
<th>HSKI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDE</td>
<td>.51**</td>
<td>.42**</td>
</tr>
<tr>
<td>CHS</td>
<td>.70**</td>
<td>.81**</td>
</tr>
<tr>
<td>ERSEA</td>
<td>.49**</td>
<td>.33**</td>
</tr>
<tr>
<td>FCE</td>
<td>.30**</td>
<td>.22**</td>
</tr>
<tr>
<td>FIS</td>
<td>.50**</td>
<td>.14**</td>
</tr>
<tr>
<td>GOV</td>
<td>.57**</td>
<td>.37**</td>
</tr>
<tr>
<td>SYS</td>
<td>.78**</td>
<td>.72**</td>
</tr>
</tbody>
</table>

TOT = Total Compliance with all Compliance Measures.
HSKI = Total Compliance with the Head Start Key Indicators.
Appendix 7 – Figure 2 – DMLMA Potential Impact on Tri-Annual Head Start Program Reviews

Present Head Start Monitoring System:

All programs receive the same Tri-Annual Reviews regardless of Compliance History:

Proposed DMLMA System with Key Indicators (KI):

100% Compliance with the Head Start Key Indicators (HSKI):

If less than 100% with the Head Start Key Indicators (HSKI):
The above proposed change is cost neutral by re-allocating monitoring staff from doing only Tri-Annual Reviews on every program to doing abbreviated monitoring via the HSKI on the highly compliant programs with periodic comprehensive full monitoring less frequently (this would change if a program did not continue to be 100% in-compliance with the HSKI), and only doing more comprehensive full monitoring on those programs with low compliance with the Compliance Measures and/or less than 100% compliance with the HSKI. Once a program was in the high compliance group they would be eligible for the HSKI abbreviated monitoring.

However, the real advantage in this proposed change is the increased frequency of targeted or differential monitoring of all programs.

**DMLMA Algorithm with Key Indicators applied to Head Start Tri-Annual Reviews:**

Six (6) Years example:

**Present Head Start Monitoring System:**

(Tri-Annual Visits)(Compliance Measures)(Percent of Programs(%)) = Total Effort
(3)(131)(100) = 39300
Total Effort = **39300**

**Revised Head Start Monitoring DMLMA with Key Indicators System:**

100% Compliance with HSKI:
(Number of Monitoring Visits)(Compliance Measures)(Percent of Programs*(%)) = Total Effort
Abbreviated Monitoring Visits using Key Indicators: (6)(8)(43*) = 2064
Full, Comprehensive Monitoring Visit using all Compliance Measures: (1)(131)(43*) = 5633

Less than 100% Compliance with HSKI:
(Number of Monitoring Visits)(Compliance Measures)(Percent of Programs**(%)) = Total Effort
Full, Comprehensive Monitoring Visits using all Compliance Measures: (4)(131)(57**) = 29868

100% Compliance with HSKI + Less than 100% Compliance with HSKI = Total Effort:
Total Effort = 2064 + 5633 + 29868 = **37565**

*This was the actual percent of Head Start Programs that met the criteria of 100% compliance with HSKI in this study.

**This was the actual percent of Head Start Programs that did not meet the criteria of 100% compliance with HSKI in this study.

It would be expected that the total population of Head Start programs would have a similar percent as was found in this representative sample (43% = 100% compliance with HSKI and 57% = less than 100% compliance with HSKI). This representative sample for this study constituted approximately 25% of all Head Start programs nationally.
This short paper will present the Key Indicators as they appear in *Stepping Stones* (3rd edition). It provides the statistically predictive standards (Key Indicators) that could determine overall compliance with *Stepping Stones* (AAP, APHA, NRC, 2013) and *Caring for Our Children* (AAP, APHA, NRC, 2011) based upon the statistical methodology (Fiene & Nixon, 1985). But before delineating the Key Indicators a few definitions need to be provided to put these key indicators in perspective.

**Definitions:**

**Risk Assessment (RA)** - a differential monitoring approach that employs using only those rules, standards, or regulations that place children at greatest risk of mortality or morbidity if violations/citations occur with the specific rule, standard, or regulation. *Stepping Stones* (3rd edition) is an example of a risk assessment approach.

**Key Indicators (KI)** - a differential monitoring approach that employs using only those rules, standards, or regulations that statistically predict overall compliance with all the rules, standards, or regulations. In other words, if a program is 100% in compliance with the Key Indicators the program will also be in substantial to full compliance with all rules, standards, or regulations. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators the program will also have other areas of non-compliance with all the rules, standards, or regulations. The key indicators put forth in this paper are an example of the approach.

**Differential Monitoring (DM)** - this is a relatively new approach to determining the number of visits made to programs and what rules, standards, or regulations are reviewed during these visits. There are two measurement tools that drive differential monitoring, one is Weighted Risk Assessment tools and the other is Key Indicator checklists. Weighted Risk Assessments determine how often a program will be visited while Key Indicator checklists determine what rules, standards, or regulations will be reviewed in the program. Differential monitoring is a very powerful approach when Risk Assessment is combined with Key Indicators because a program is reviewed by the most critical rules, standards, or regulations and the most predictive rules, standards, or regulations. See Fiene’s Logic Model & Algorithm for Differential Monitoring *(DMLMA©)* (Fiene, 2013).

**Early Childhood Program Quality Indicator Model (ECPQIM)** (Fiene, 2013; Fiene & Kroh, 2000; Griffin & Fiene, 1995; Fiene & Nixon, 1985) – this definition is provided to place the results of this paper into the larger program monitoring systems perspective. ECPQIM are models that employ a key indicator or dashboard approach to program monitoring. Major program monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With these models, it is possible to compare results obtained from licensing systems, quality rating and improvement systems (QRIS), risk assessment systems, key indicator systems, technical assistance, and child development/early learning outcome systems. The various approaches to validation (Zellman & Fiene, 2012) are interposed within this model and the specific
expected correlational thresholds that should be observed amongst the key elements of the model are suggested. Key Elements of the model are the following: CI = Comprehensive Instrument - state or federal standards, usually rules or regulations that measure health and safety - Caring for Our Children or Head Start Performance Standards will be applicable here. Quality Rating and Improvement Systems (QRIS) standards at the state level; ERS (ECERS, ITERS, FDCRS), CLASS, or CDPES (Fienie & Nixon, 1985). RA = Risk assessment tools/systems in which only the most critical rules/standards are measured. Stepping Stones is an example of this approach. KI = Key indicators in which only predictor rules/standards are measured. The Thirteen Indicators of Quality Child Care (Fiene, 2003) is an example of this approach. DM = Differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol. Technical assistance/training and/or professional development system which provides targeted assistance to the program based upon the Differential Monitoring results. And finally, child outcomes which assesses how well the children are developing which is the ultimate goal of the system.

The Key Indicators from Stepping Stones (3rd Edition)¹

1.1.1.2 - Ratios for Large Family Child Care Homes and Centers
   1.3.1.1 - General Qualifications of Directors
   1.3.2.2 - Qualifications of Lead Teachers and Teachers
   1.4.3.1 - First Aid and CPR Training for Staff
   1.4.5.2 - Child Abuse and Neglect Education
   2.2.0.1 - Methods of Supervision of Children
     3.2.1.4 - Diaper Changing Procedure
     3.2.2.2 - Handwashing Procedure
     3.4.3.1 - Emergency Procedures
   3.4.4.1 - Recognizing and Reporting Suspected Child Abuse, Neglect, and Exploitation
   3.6.3.1 - Medication Administration
   5.2.7.6 - Storage and Disposal of Infectious and Toxic Wastes
   6.2.3.1 - Prohibited Surfaces for Placing Climbing Equipment
   7.2.0.2 - Unimmunized Children
   9.2.4.5 - Emergency and Evacuation Drills/Exercises Policy
Just as there has been three editions of *Caring for Our Children* and *Stepping Stones*, this paper and the resulting Key Indicators represents the third edition of Key Indicators for early care and education. The first two editions are represented in the publications by Fiene & Nixon (1985) and Fiene (2003) respectively (see the reference list below).

**References**


**Notes:**

1. Please see *Stepping Stones (3rd edition)* and *Caring for Our Children (3rd edition)* for the details of each Key Indicator.

2. For the reader who is interested in learning more about the DMLMA/ECPQIM model, please refer to these publications which are available through the following website:

   [http://RIKInstitute.wikispaces.com](http://RIKInstitute.wikispaces.com)

**For additional information, please contact:**

Richard Fiene, Ph.D., Director
Research Institute for Key Indicators
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
RIKI.Institute@gmail.com
717-944-5868 Phone and Fax
[http://RIKInstitute.wikispaces.com](http://RIKInstitute.wikispaces.com)
[http://pennstate.academia.edu/RickFiene](http://pennstate.academia.edu/RickFiene)