

The Relationship of Licensing, Head Start, Pre-K, QRIS, Accreditation, and Professional Development and their Potential Impact on Child Outcomes

Richard Fiene, Ph.D.

October 11, 2013

ABSTRACT

This short paper will provide some thoughts about the various public policy initiatives/systems to improve early care and education, such as licensing, Head Start, Pre-K, QRIS, accreditation, and professional development and their potential impact on child outcomes. Early care and education is at a major crossroads as a profession in attempting to determine which quality initiatives have the greatest impact on children. Results are starting to come in from early studies which may provide some guidance as policy makers begin making decisions about where to focus their limited funding resources.

Improving early care and education programs has a long public policy history as we attempt to find the most cost effective and efficient means for attaining this lofty goal. There have been many ups and downs over the years where funding was adequate and when it was not, but our desire to accomplish this goal has always been front and center. Now, as a profession, we are at somewhat of a cross-roads in determining which of the many quality initiatives appear to have the greatest impact on children's development. When I refer to children's development, I am looking at the whole child from the perspective of a child's developmental status as well as the child's health and safety.

Presently we have many quality initiatives to look at which is a very good thing since at times in the past we did not always have so many choices. Probably the one constant throughout the history of early care and education in the past century has been licensing or regulations/rule formulation. Some many argue that licensing is not a quality initiative but I would suggest that licensing has many of the structural aspects of quality that have been identified in the research literature. The other quality initiatives I will discuss have really started and been implemented in the very later part of the 20th century so we are talking about a relatively new science when we think about having its intended impact on children. Also, I am talking about large public policy initiatives rather than highly structured, single focused research studies involving small samples of children.

Let's start with licensing since this system has been present for the longest period of time. The purpose of licensing is to act as the gatekeeper to the early care and education field in which only those providers who meet specific standards, generally called rules or regulations are permitted to operate and care for children. The rules are dominated by health and safety concerns with less emphasis on curriculum planning and staff-child interactions. The rules measure more structural aspects of quality than the process aspects of quality; dealing with what attorney's call the "hard data" rather than the "soft data".

Since licensing rules allow entry into the early care and education field to provide services usually the rules are not overall stringent with the majority of providers being in high compliance if not full compliance with all the rules. This would be expected since these are basic health and safety standards. And in fact when one looks at compliance data, it is extremely skewed with the majority of providers having very high compliance scores with relatively few violations of the rules. However, this does introduce a certain difficulty in using these data for decision making purposes at an aggregate level because so many providers score at a high level it becomes increasingly difficult to distinguish between the really excellent providers and the somewhat mediocre providers. Another way of looking at this skewing of the data is to term it as a plateau effect in which there is very little variance at the upper ends of the compliance spectrum. This is a major issue with skewed data and basic standards which is an important consideration with licensing but will also be an important consideration when one looks at the other quality initiatives to be addressed shortly.

Because of this plateau effect with licensing data, it may explain much of the lack of relationships found between compliance with rules and any types of outcomes related to children's outcomes and provider's overall quality. However, with licensing data and making comparisons to children's outcomes we should be looking at general health data such as immunization status and safety data such as the number of injuries at programs with varying levels of compliance with health and safety rules.

A significant development over the past two decades has been the development of national health and safety standards with the publication of Caring for Our Children (CFOC3) and Stepping Stones (SS3). Although these standards are not required but are only recommended practice that provides guidance to states as they revise their rules, these two documents have been embraced by the licensing/regulatory administration field. Although unlikely, if not impossible, to comply with all the CFOC3 standards, it would be interesting to compare states on this set of standards which may add a good deal of variance to the basic health and safety data that has been missing with licensing rules.

The next system to look at is the national Head Start program. Out of the major programs that are national in scope, Head Start has a long history of providing services to low income children and their families. Head Start Performance Standards are definitely more stringent than licensing rules but not as stringent as accreditation standards. Based upon Head Start's more stringent

standards and the additional supports that are part of its program, Head Start generally scores higher on program quality tools (e.g., CLASS or ERS) than licensed child care in states.

With Head Start programs, we at times find skewing or plateauing of data when we compare compliance with the Head Start Performance Standards (HSPS) and program quality tools such as the CLASS. However, this is dependent upon the various subscales within the CLASS in which the plateauing of data does not occur all of the time. I think that has a lot to do with the HSPS being fairly stringent standards as compared to state licensing rules in general.

A program that has gotten a good deal of support at the state level are Pre-K programs. These programs come with stricter standards than licensed child care with an emphasis on the professional development of staff. There is more concern about the process aspects of quality which focus more on teacher-child interactions. This emphasis on teacher-child interaction has paid off in which these programs generally are high performers when you compare Pre-K funded classrooms to licensed child care classrooms. In fact, Pre-K funding appears to have a positive impact on licensed child care in raising overall quality scores on the ECERS-R for all classrooms in programs that receive Pre-K funding even if some of the classrooms are not the direct beneficiaries of the funding. This is a very significant finding because we knew that Pre-K funding increased the quality of care in classrooms receiving those funds, but now, it appears that there is a spillover effect to all classrooms co-located with Pre-K funded classrooms. I must admit that I was initially skeptical when Pre-K funding was first proposed because I thought it would take funding and the focus away from improving licensed child care at the state level; but it appears that the advocates for Pre-K were right in their assertion that Pre-K would increase the quality of all early care and education which includes licensed child care.

A more recent entry into the state funding scene are QRIS (Quality Rating and Improvement Systems) which build upon licensing systems, are voluntary, and have substantial financial incentives for participating in this quality improvement system. It is too early to really determine if QRIS is having the intended impact because the program is so new (50% of states have a QRIS), and the penetration rate is usually below 50% in any given state (remember the system is voluntary). However, in the few studies done, the results are mixed. It does appear that programs which move up the various star levels do increase the quality of care they provide; but in a most recent study looking at child outcomes, no relationship was found between increasing levels of compliance with QRIS standards and how well children did in those programs with the exception of CLASS scores in which teacher-child interactions were measured and emphasized – here there were significant relationships between higher scores on the CLASS and child outcomes.

Accreditation systems come in many varieties but there are only three that I know of in which empirical studies have been done to validate their systems: NAEYC, NECPA for centers and NAFDC for homes. Also reliability testing has been done in each of these systems.

Accreditation is a rigorous self-study that really improves programs through the self-study

process. This should come as no surprise because we have known for some time that program monitoring all by itself leads to program improvements. Now when you couple that with technical assistance you see even more improvement. Accreditation is usually the other pillar of a QRIS system with licensing being the first pillar. The QRIS standards fill the gap from licensing to accreditation. Accreditation is a voluntary system just as in most cases with QRIS. However, in accreditation we are reaching less than 10% of the programs with the majority of these attaining NAEYC accreditation. NECPA and NAFDC have much smaller market shares.

The last system to be addressed is the professional development systems that have been established in all states. This is one quality improvement initiative that has 100% penetration in all states. It is usually tied to QRIS through technical assistance and mentoring (coaching). When it focuses on mentoring rather than workshops, it has demonstrated its effectiveness in changing teachers behaviors in how they interact with children in their care in a very positive fashion. This is very important because the research literature is clear about the importance of the teacher-child interaction when it comes to child outcomes. Professional development runs the gamut from pre-service (University based programs) to in-service (training, technical assistance, mentoring, coaching) programming for teachers and directors.

So where does this leave us when policy makers begin to try to determine which quality improvement initiatives should be invested in to start with, which to increase in funding, and maybe even which ones should be defunded. I think there are some trends we need to begin to look at, such as the following:

- 1) Having stringent and rigorous standards is very important. The more that we do not, the more opportunities for mediocre programs to score artificially higher on whatever scale that is used. This is evident with licensing data where the data are significantly skewed with a major plateau effect at the upper end of compliance rules/regulations.
- 2) Emphasis on teacher-child interaction needs to be paramount in our quality improvement initiatives. Working with teachers through mentoring/coaching appears to be most effective in changing teachers' behaviors in interacting more positively with children.
- 3) Making sure we are measuring the right outcomes. Match health and safety standards with health and safety outcomes for children. Match developmental outcomes for children with standards that emphasize positive teacher-child interactions.
- 4) Building upon #1 above, find what the key indicators are with all the data that we collect. We are spending too much time in looking at too many things which in many cases are simply just not the right things to look at. As states' data systems become more sophisticated, and they are, this will be easier to do. Let's begin to utilize the data we have already collected.

DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (DMLMA)[®]: A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL⁴ (ECPQIM⁴)[®] FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES

Richard Fiene, Ph.D.

This Differential Monitoring Logic Model and Algorithm (DMLMA[®]) 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 Table 1 and Figures 1 & 2).

The DMLMA[®] can be used by state agencies (child care, child residential, adult residential (just replace Child Outcomes with Adult Outcomes)), Federal agencies (Head Start, child care, Pre-K), and large provider organizations where an economy of scale is required. This model can be used with state as well as national standards, such as state licensing rules/regulations and *Head Start Performance Standards* or *Caring for Our Children/Stepping Stones*. Most states and Federal agencies have either some or all of the key elements of this model in their overall monitoring systems. The purpose of this model is to alter a one-size fits all monitoring system to one that is targeted, spending more time with problem programs who need additional assistance. This is a cost neutral model that is both cost effective and efficient and re-allocates resources from the compliant programs to the non-compliant programs.

Key Elements (see Figures 1 & 2): **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 (Figure 2) of the overall system and assessing the inter-correlations (Figure 1) to determine that the DM system is improving the health, safety, program quality and ultimately the overall development of the children it serves.

The DMLMA[®] provides a cross-cutting methodology that can be used in all early care and education delivery systems as well as in other human services. In the past many of these monitoring systems have functioned in silos. The DMLMA[®] integrates all these various monitoring systems together so that the overall monitoring system can be validated as being cost effective and efficient.

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 Table 1).

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.

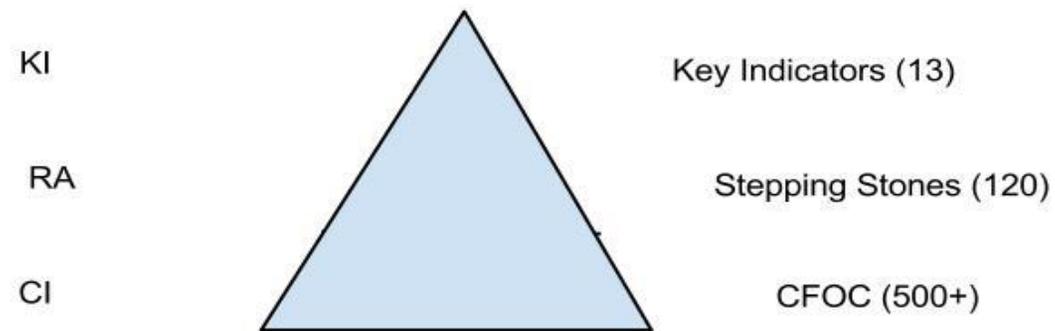
The **sixth step** is for the state to generate a Key Indicator (KI) tool from the CI data base. Please see Fiene & Nixon (1985) and Fiene & Kroh (2000) for a detailed explanation of the methodology for generating a KI tool. This step is also part of the second approach to validation which involves Measures. 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 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 children's development other than child care facilities.

Validation is a continuous approach and is not a once and done process. States should look at their monitoring systems on an on-going basis and make the necessary adjustments as data are collected and compared in order to keep program monitoring as cost effective and efficient.

Relationship of Key Indicators (KI), Stepping Stones (RA), and Caring for Our Children (CFOC)(CI)



The above diagram depicts the relationship amongst KI, RA, and CI in which the full set of rules is represented by CFOC - Caring for Our Children, followed by RA which are the most critical rules represented by Stepping Stones, and finally the predictive rules represented by the 13 Key Quality Indicators.

Table 1: DMLMA[®] Expected Thresholds

Key Elements	PQ	RA	KI	DM	PD	CO
CI	0.3	0.5	0.7	0.5	0.5	0.3
PQ				0.3	0.3	0.3
RA			0.5	0.5	0.5	0.3
KI				0.5	0.5	0.3
DM					0.5	
PD						0.3

RELATED PUBLICATIONS:

Fiene (2007). Child Development Program Evaluation & Caregiver Observation Scale, in T Halle (Ed.), *Early Care and Education Quality Measures Compendium*, Washington, D.C.: Child Trends.

Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, Winter 2002-2003, pps 12-13.

Fiene (2002). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.

Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services*, 5(2), 38-39.

Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice*, 1(2), 10-11.

Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.

Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, 14(3), 198-214.

Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.

Morgan, Stevenson, Fiene, & Stephens (1986). Gaps and excesses in the regulation of child day care, *Reviews of Infectious Diseases--Infectious Diseases in Child Day Care: Management and Prevention*, 8(4), 634-643.

Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

For additional information, please contact:

Dr Richard Fiene, Director, Research Institute for Key Indicators, DrFiene@gmail.com; ResearchInstituteKeyIndicators@ymail.com

Figure 1: Differential Monitoring Logic Model & Algorithm (DMLMA)[®] Thresholds

DMLMA[®] Expected Thresholds:
High Correlations (.70+) = CI x KI.
Moderate Correlations (.50+) = CI x RA; RA x DM; RA x KI; KI x DM; DM x PD.
Lower Correlations (.30+) = PQ x CI; PQ x CO; PQ x DM; RA x CO; KI x CO; CI x CO.

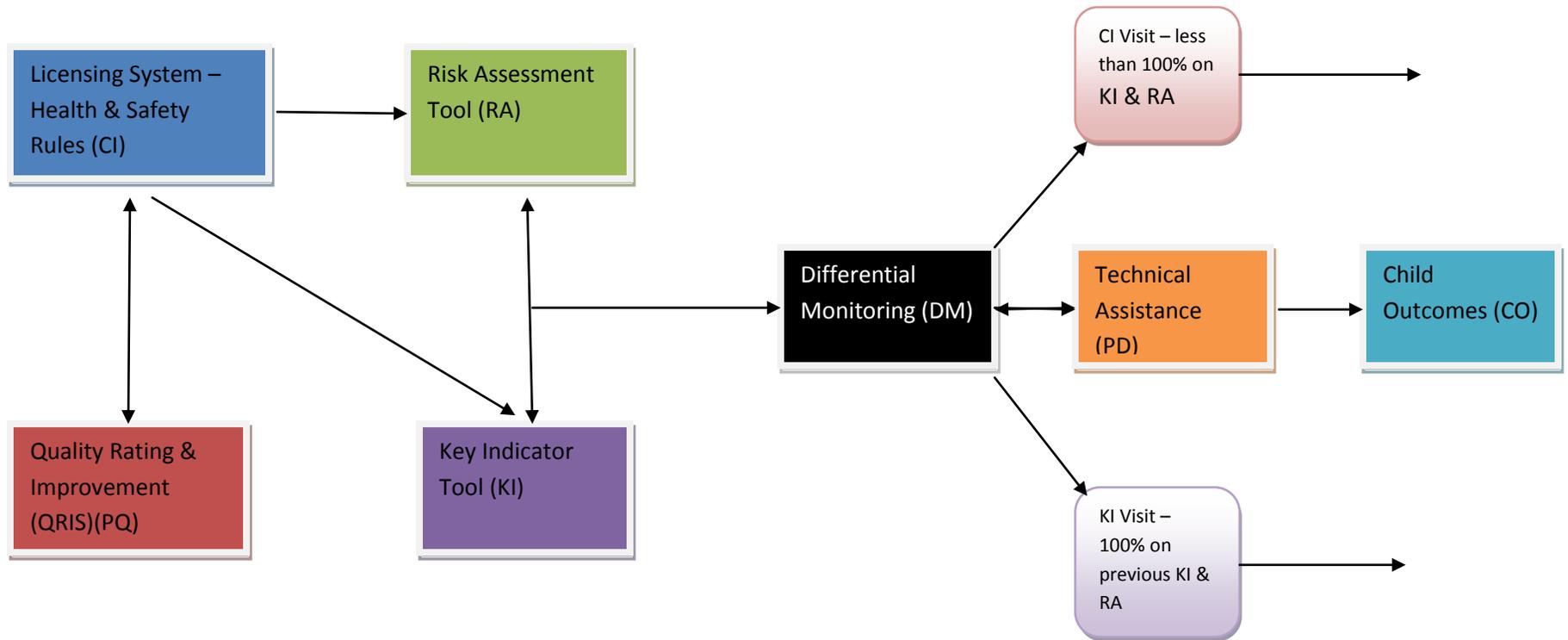
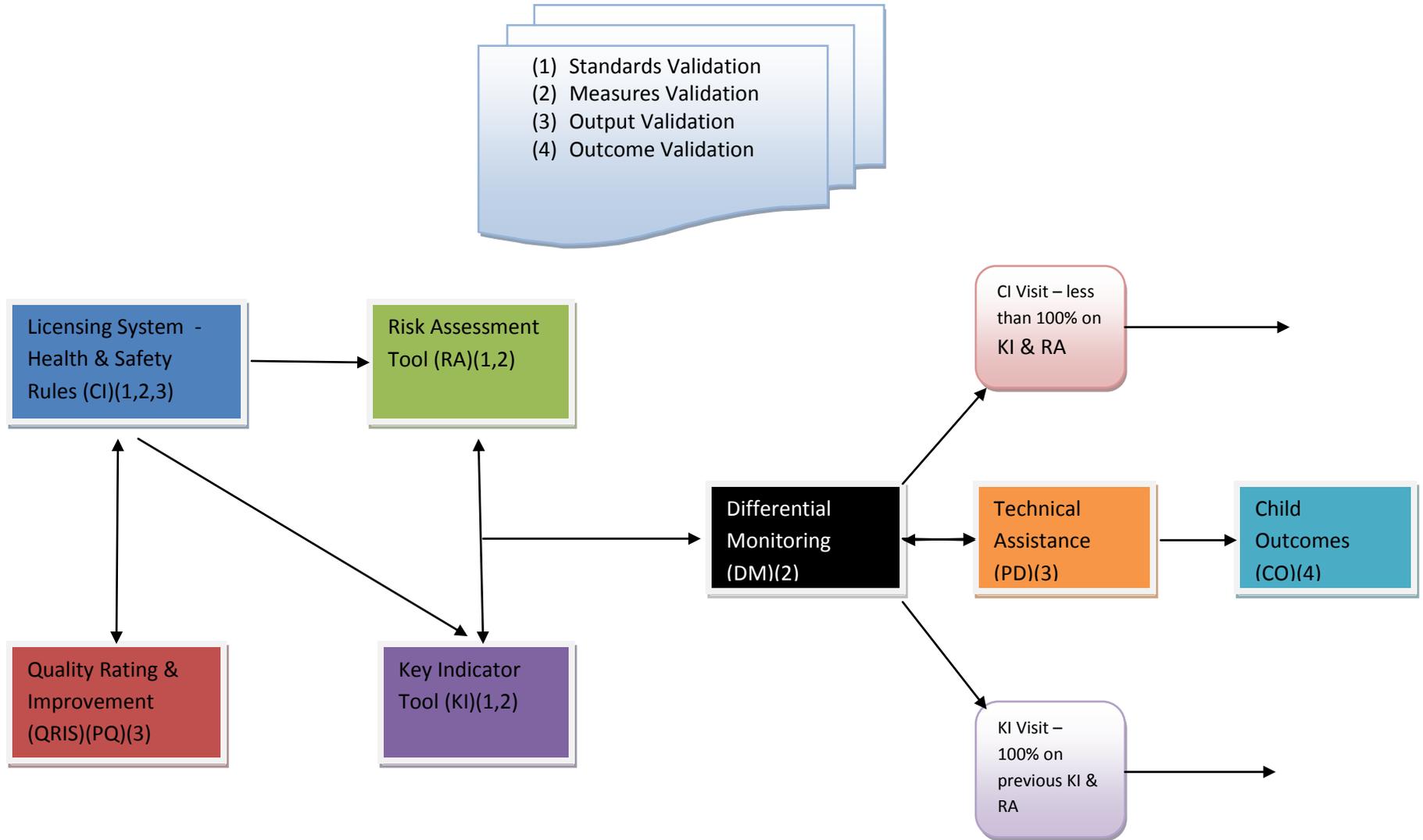


Figure 2: Differential Monitoring Logic Model & Algorithm (DMLMA)[®] and Validation Approaches (Zellman & Fiene, 2012)

$$\sum CI \times \sum PQ \Rightarrow \sum RA + \sum KI \Rightarrow \sum DM + \sum PD \Rightarrow CO$$



DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL (ECPQIM⁴©) FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES

Richard Fiene, Ph.D.

December 2014

This Differential Monitoring Logic Model (DMLM©) provides a new Early Childhood Program Quality Indicator Model (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 (see Figures 1 & 2 for a graphical depiction of the theoretical underpinnings and actual design & logic model for the ECPQIM⁴©/DMLM).

The DMLM© can be used by state agencies (child care, child residential, adult residential (just replace Child Outcomes with Adult Outcomes)), Federal agencies (Head Start, child care, Pre-K), and large provider organizations where an economy of scale is required. This model can be used with state as well as national standards, such as state licensing rules/regulations and *Head Start Performance Standards* or *Caring for Our Children/Stepping Stones*. Most states and Federal agencies have either some or all of the key elements of this model in their overall monitoring systems. The purpose of this model is to alter a one-size fits all monitoring system to one that is targeted, spending more time with problem programs who need additional assistance. This is a cost neutral model that is both cost effective and efficient and re-allocates resources from the compliant programs to the non-compliant programs.

Key Elements (see Figure 2): **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; in other words, to determine if the DM system is improving the health, safety, program quality and ultimately the overall development of the children it serves.

The DMLM© provides a cross-cutting methodology that can be used in all early care and education delivery systems as well as in other human services. In the past many of these monitoring systems have functioned in silos. The DMLM© integrates all these various monitoring systems together so that the overall monitoring system can be validated as being cost effective and efficient.

STATE AGENCY PLAN for implementing a Differential Monitoring System:

The **first step** in utilizing the DMLM© 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 DMLM© 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.

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.

The **fourth step** for the state is to compare the results from the CI with the RA tools.

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.

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. 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. Also, this step should drive decisions within the technical assistance/training/professional development (PD) system in what resources are allocated to a particular facility.

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/Systems and Standards (ELS) are developed, this will become more common place.

The ECPQIM⁴©DMLM© is presented without two additional items that were present in the 2012/2013 versions which are important to note. The algorithm (Fiene, 2012, 1013) and validation framework (Zellman & Fiene, 2012) are not presented because the author felt that these two components took away from a more direct presentation of differential monitoring. For those interested readers, please refer to my previous abstracts (Fiene, 2012, 2013) which included the algorithm and validation frameworks.

Just a brief word about the Theoretical Underpinnings for ECPQIM⁴. This graphic (Figure 1) attempts to provide the relationships amongst public policy, interventions, and empirical evidence through the lens of translational research, implementation science, and program monitoring. In constructing the ECPQIM⁴ I have borrowed concepts from each area and integrated them in a model for monitoring early care and education programs. The graphic provides a means for displaying the relationships and potential intersections as well as the content that is important to each scientific/research field.

Figure 3 is provided as additional information regarding differential monitoring conceptually without all the details as in figure 2; and figure 4 is provided to demonstrate the impact that a state's licensing law can have on using the Key Indicators and Risk Assessment methodologies. Also, taking Figure 2 and attempting to quantify these relationships, I am proposing the following scoring protocol as depicted in Table 1. This can provide a numerical means of comparing various differential monitoring systems and their relative strength. This protocol could be a useful tool in future research for determining which combinations work best.

Table 1: Differential Monitoring Scoring Protocol© (Fiene, 2014)

Score	Systems Present
0	No systems in place.
2	KI or RA in place and not linked.
4	(KI & RA in place but not linked) or (PC + PQ are linked).
6	(KI & RA in place) & (KI + RA are linked)
8	(KI & RA in place but not linked) & ((PC + PQ) are linked).
10	All systems in place and linked.

RELATED PUBLICATIONS:

Fiene (2012/2013). Early childhood program quality indicator model (ECPQIM⁴©): Differential monitoring logic model and algorithm (DMLMA©), Middletown, PA: Research Institute for Key Indicators LLC (RIKI).

Fiene (2007). Child Development Program Evaluation & Caregiver Observation Scale, in T Halle (Ed.), *Early Care and Education Quality Measures Compendium*, Washington, D.C.: Child Trends.

Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, Winter 2002-2003, pps 12-13.

Fiene (2002). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.

Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services*, 5(2), 38-39.

Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice*, 1(2), 10-11.

Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.

Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, 14(3), 198-214.

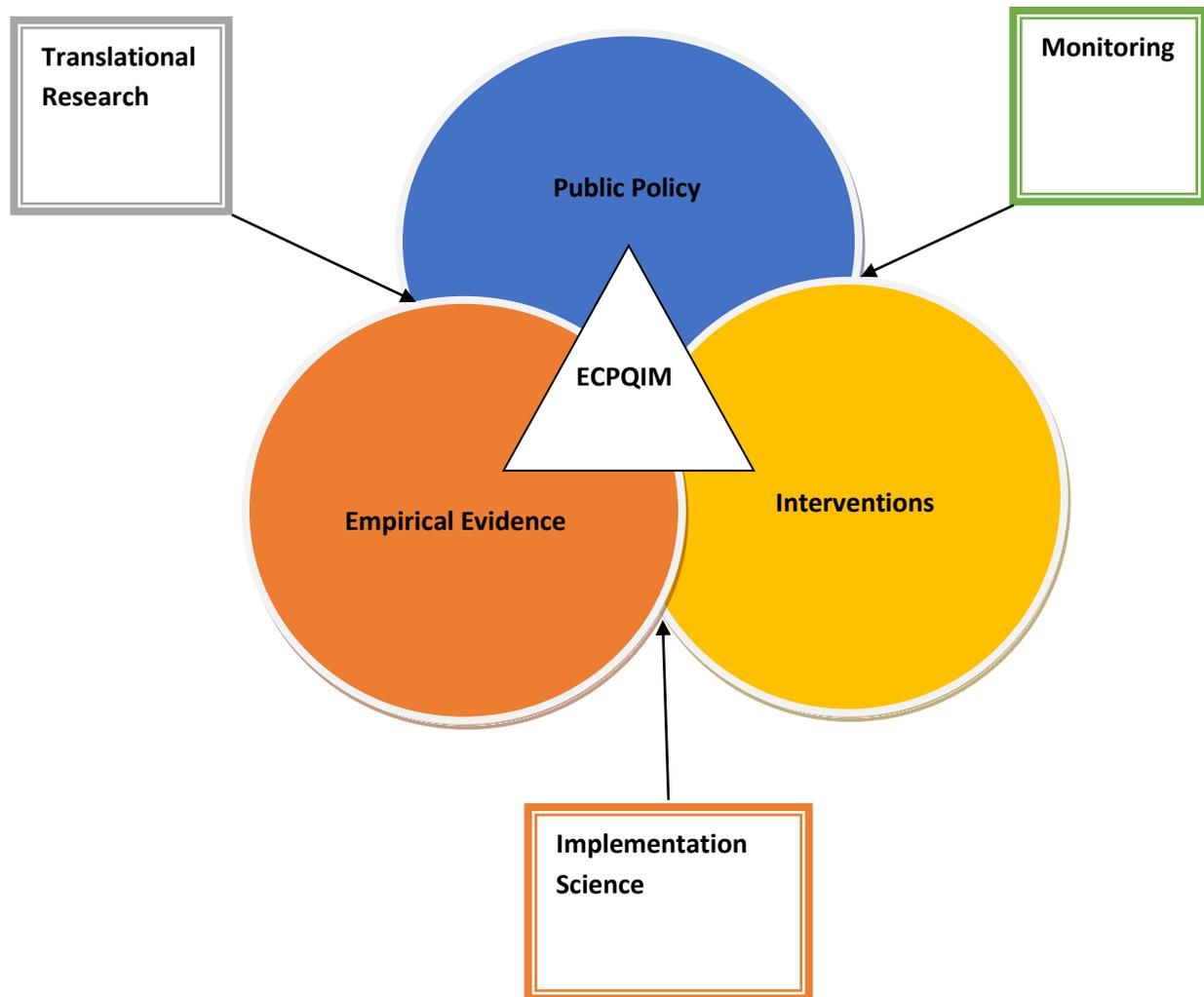
Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.

Morgan, Stevenson, Fiene, & Stephens (1986). Gaps and excesses in the regulation of child day care, *Reviews of Infectious Diseases--Infectious Diseases in Child Day Care: Management and Prevention*, 8(4), 634-643.

Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Figure 1

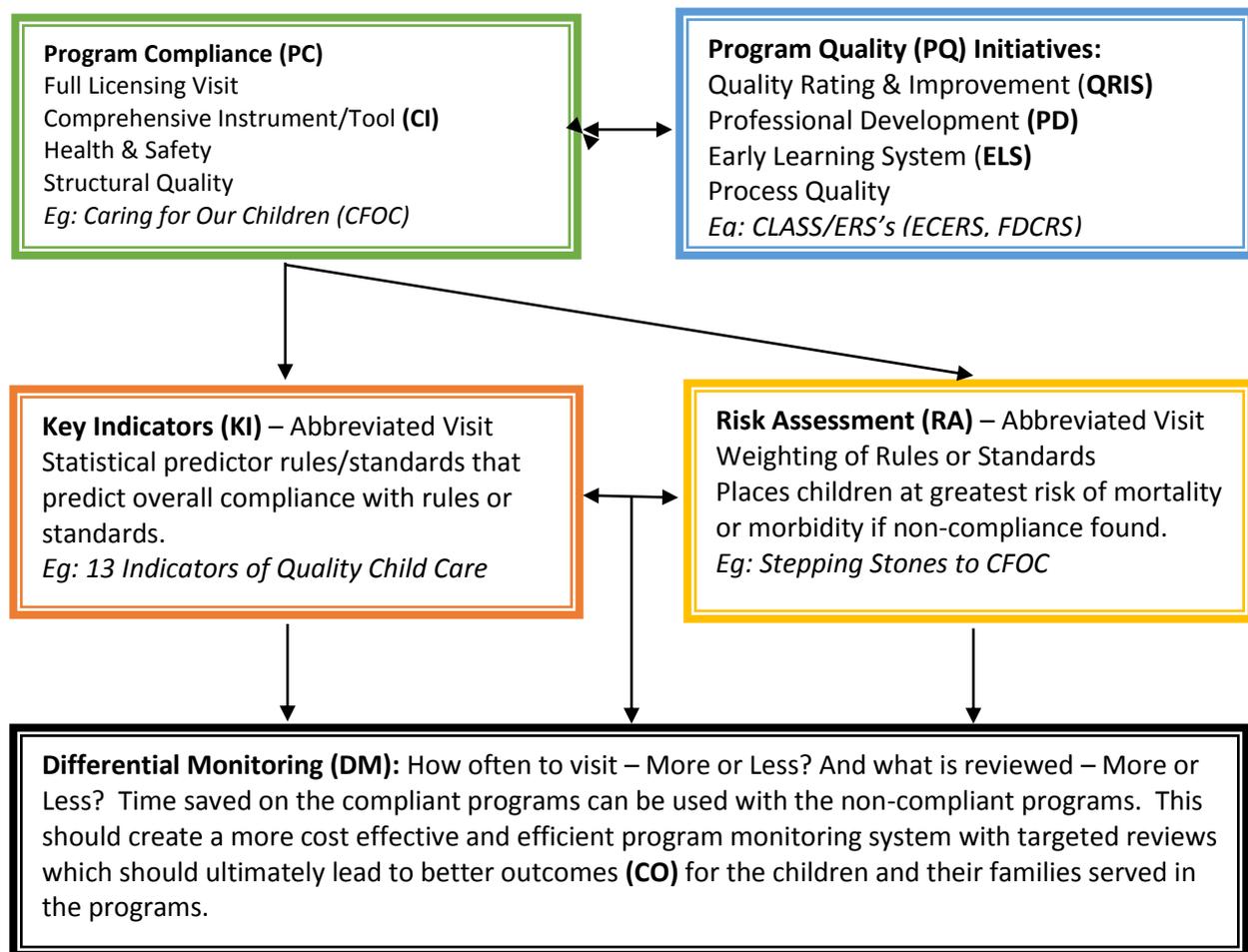
The Scientific Underpinnings for ECPQIM⁴: Early Childhood Program Quality Indicator Model©



ECPQIM⁴© Fiene (2014)

Figure 2

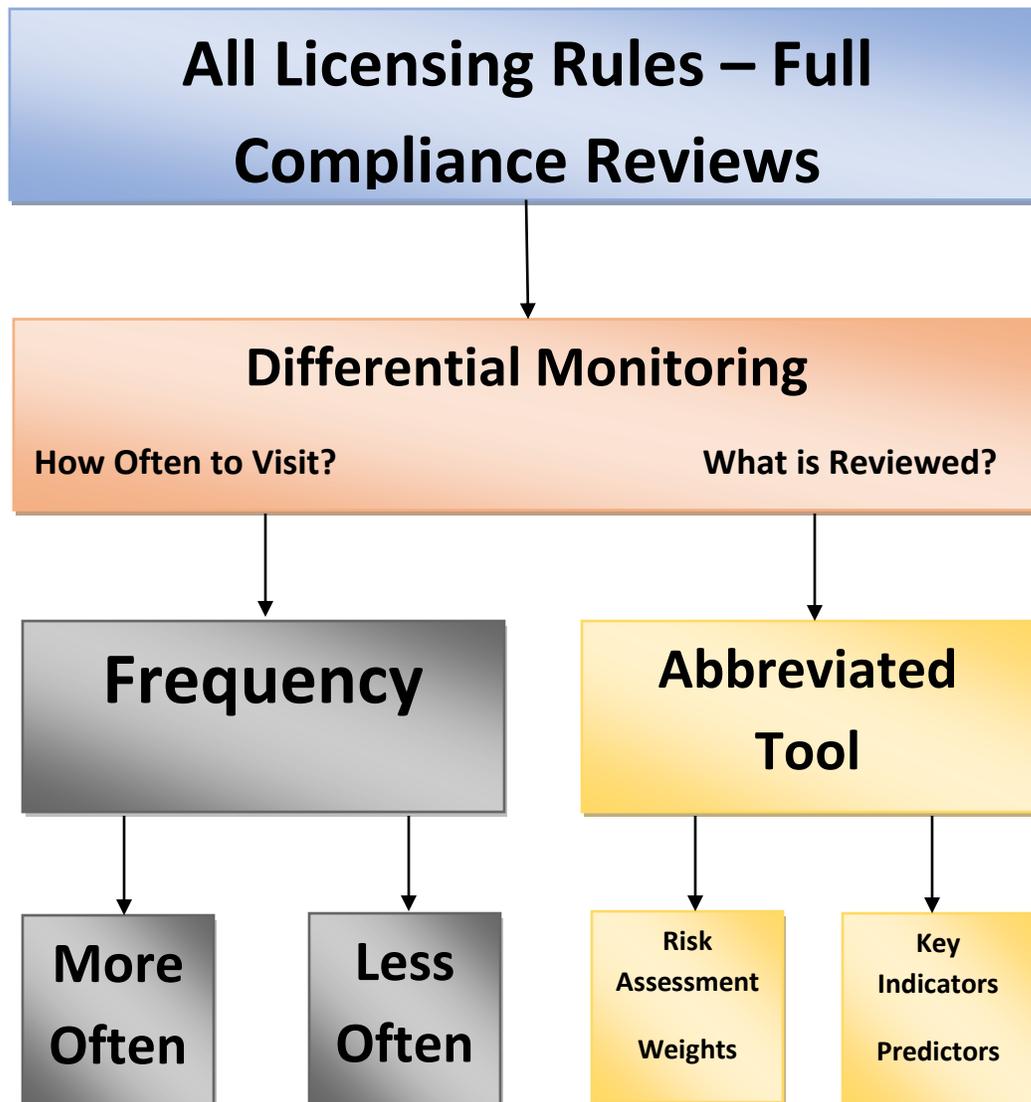
**Early Childhood Program Quality Indicator Model (ECPQIM⁴©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Comprehensive National Example**



DMLM© Fiene (2014)

Figure 3

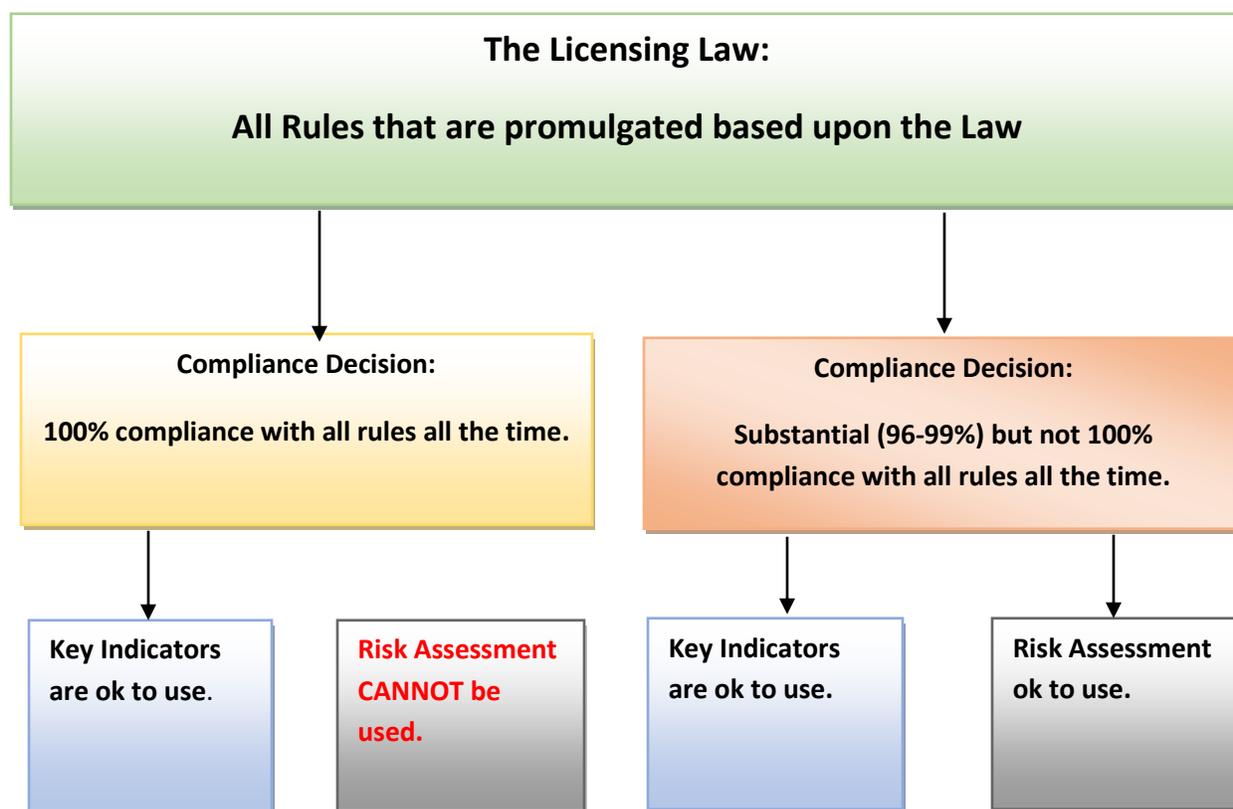
Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators



Fiene (2014). DMLMA/ECPQIM⁴©, RIKI.

Figure 4

When Key Indicators and Risk Assessments Can Be Used



Fiene (2014)

Early Childhood Program Quality Indicator Model (ECPQIM⁴©)/Differential Monitoring Logic Model (DMLM©) Comprehensive National and State Examples of Its Application

Richard Fiene, Ph.D.

December 2014

I have written extensively about an Early Childhood Program Quality Indicator Model (ECPQIM⁴©) and its latest iteration presented as a logic model: Differential Monitoring Logic Model (DMLM©). Several states and Head Start have used the model in order to re-align their program monitoring systems. This paper will present the results of those new program monitoring systems through the lenses of the ECPQIM⁴©/DMLM© logic model display. Each particular approach used various components of the overall comprehensive national model and I have highlighted those sections in highlight **yellow** and connected by an arrow.

The interested reader should obtain a copy of the Office of Child Care's *Licensing Brief on Differential Monitoring, Risk Assessment, and Key Indicators* published by the National Center on Child Care Quality Improvements which gives additional details regarding these approaches and methodologies as well as other state examples. Please go to the following URL website: (https://childcareta.acf.hhs.gov/sites/default/files/1408_differential_monitoring_final_1.pdf). In fact, this paper builds upon that excellent *Licensing Brief*.

Let's start with Figure 1 which provides the Comprehensive National Example that depicts all the possible interconnections and gives national examples from the research literature. As one will see, it is possible for a national organization or a state agency to select the various components from the model based upon what is available in their particular organization or state. All do have the program compliance/licensing component but not all have fully functional program quality initiatives or do not have the data to draw from the program quality initiatives.

The next level of components are the key indicator and risk assessment approaches or methodologies which organizations or state agencies can use alone or in tandem. One limitation in the key indicator methodology is not to use it with program initiatives if the data are not severely skewed in their data distribution as is the case with licensing data.

The last component is the resulting differential monitoring approach based upon the results from using the key indicator and risk assessment methodologies either alone or in tandem. This is the ultimate revision of the program monitoring system in which how often and what is reviewed are answered.

All the components are highlighted in Figure 1 because all are possibilities to be used by a national or state agency. The examples in Figure 1 are drawn from the national research literature so *Caring for Our Children* is the example for Program Compliance, Licensing, Health & Safety Comprehensive Instrument. The following examples in Figures 2-7 will show some differences in how national and state agencies have developed their respective differential monitoring systems. The tables (Tables 1-3) at the end of this paper (page 10) explains the scoring protocol. Also see the end of the paper for an explanation of Notes a,b,c (page 9).

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Comprehensive National Example (Maximum of 10 Points)
Figure 1**

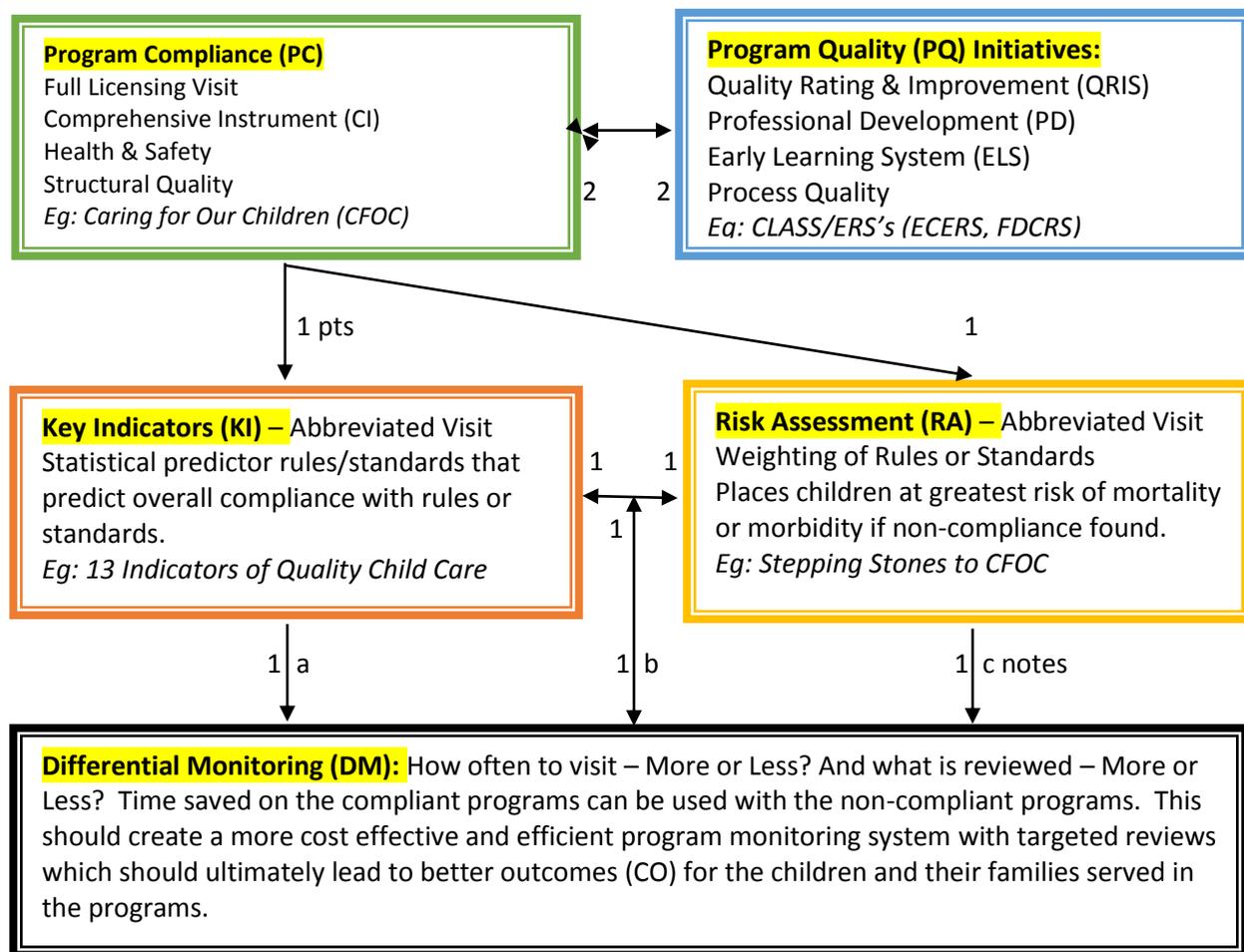


Figure 2 provides an example from New York where the state agency is attempting to restructure their early care and education program monitoring system to have a better balance between licensing and key program quality indicators. The plan is to have licensing staff collect data from both areas which means a need to save time in the licensing reviews via key indicators and to only identify indicators of quality through a risk assessment approach. The results from these two methodologies will then be combined into a Quality Indicators Instrument to be used by licensing staff in their annual reviews.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
NYQI Example (NY)
Figure 2**

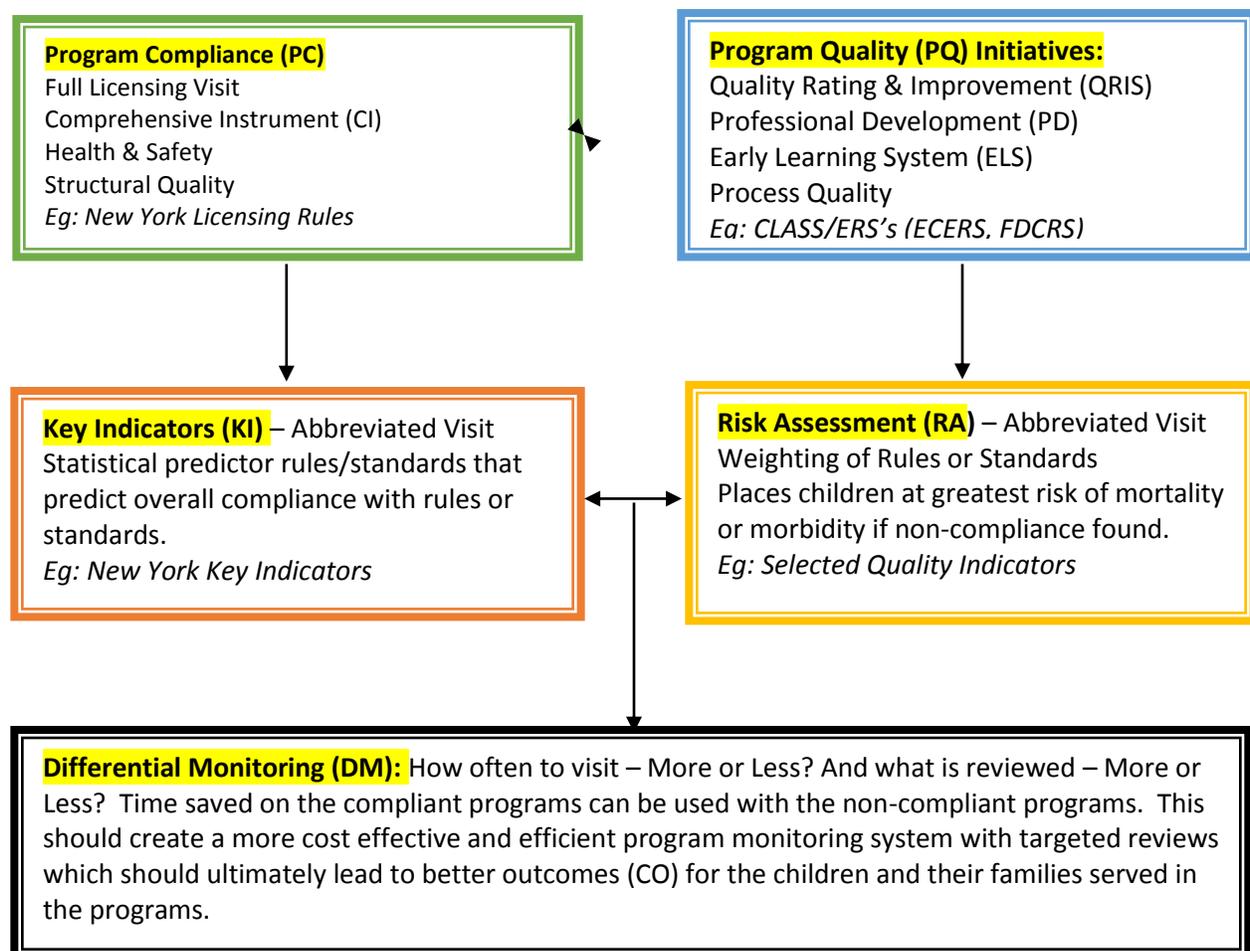


Figure 3 provides an example from Georgia in which the driving methodology is a risk assessment core rule review system that results in a differential monitoring system called the Annual Compliance Determination Worksheet (ACDW) approach. Key indicators are not used directly but were used as part of the risk assessment core rule development. Please note how the relationship amongst the various components is different from the NYQI approach delineated in Figure 2. There is a link to their program quality initiatives which proved very significant in the validation studies performed on their Core Rule differential monitoring system.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Georgia Example (GA)
Figure 3**

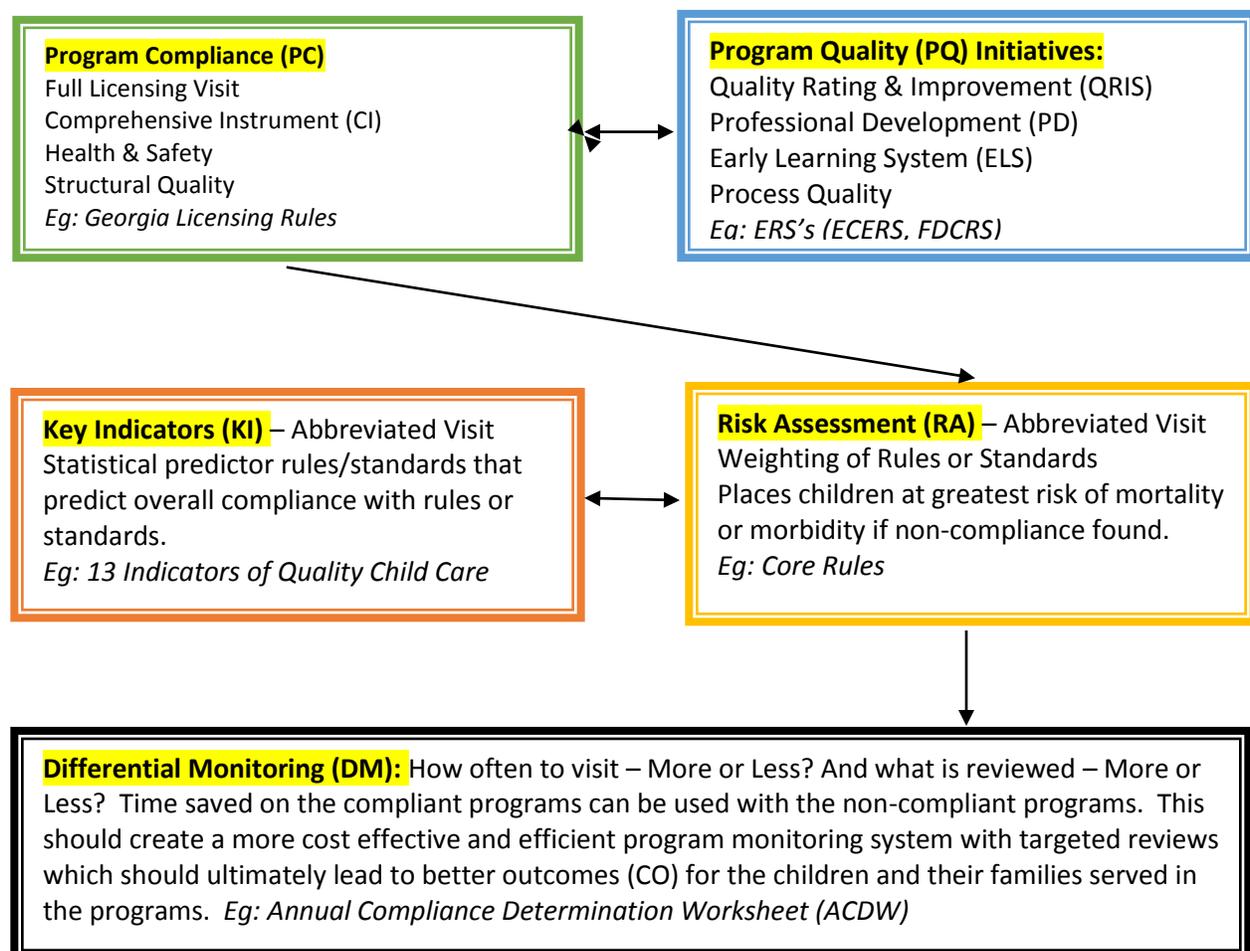


Figure 4 presents a very different approach from the previous two approaches. In Kansas's case, the state agency was only interested in developing a key indicator approach and was not interested in risk assessment nor had the capability to tie data together from their program quality initiatives. This is noted by the arrow connections and yellow highlighting which is more minimal in this figure. As one can see, this still is a viable option for developing a differential monitoring approach.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Kansas Example (KS)
Figure 4**

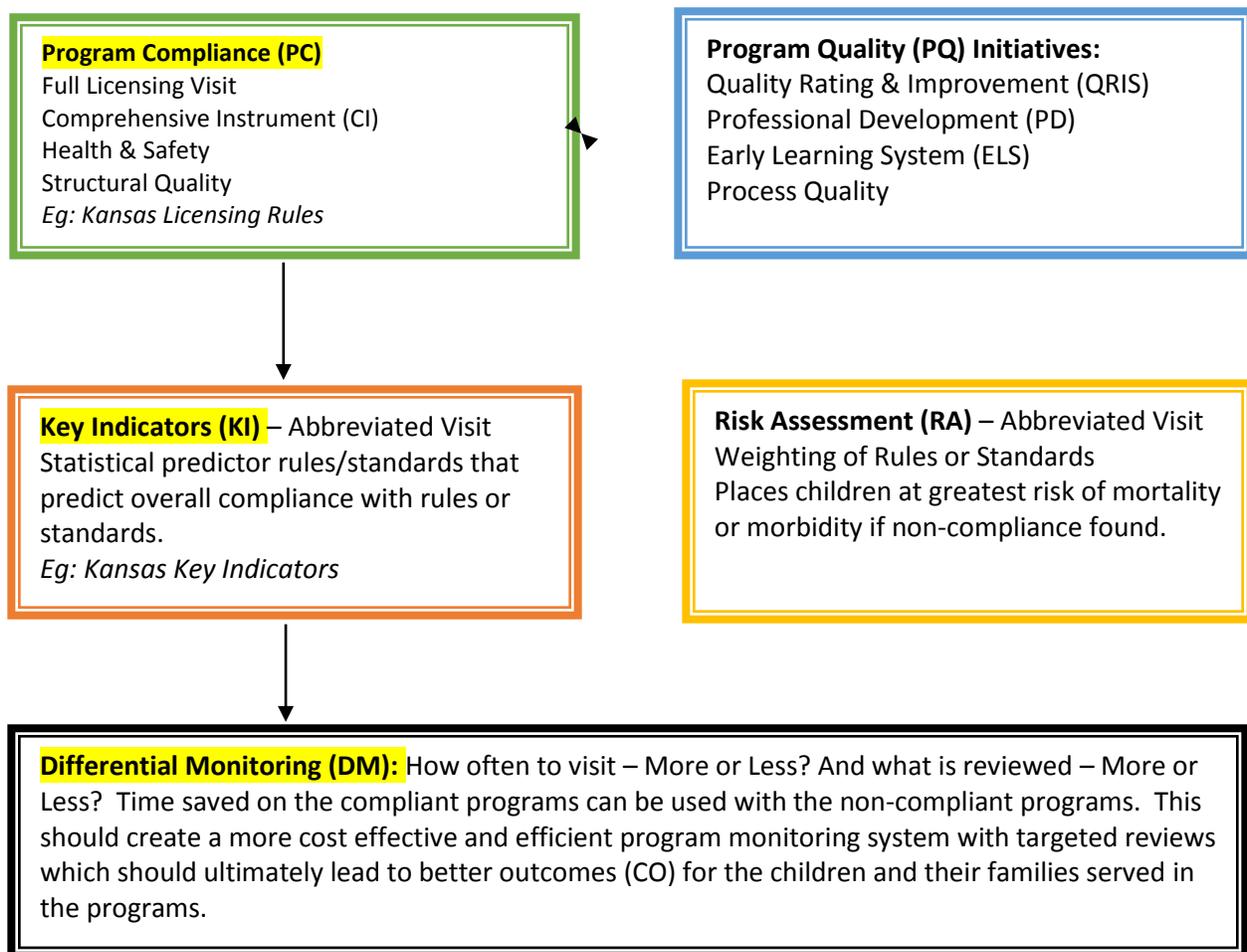


Figure 5 depicts the use of both key indicator and risk assessment methodologies with their licensing system but no data interaction with their program quality initiatives. It is proposed that both methodologies will be used together in future licensing reviews of programs which will constitute their differential monitoring system approach.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Illinois Example (IL)
Figure 5**

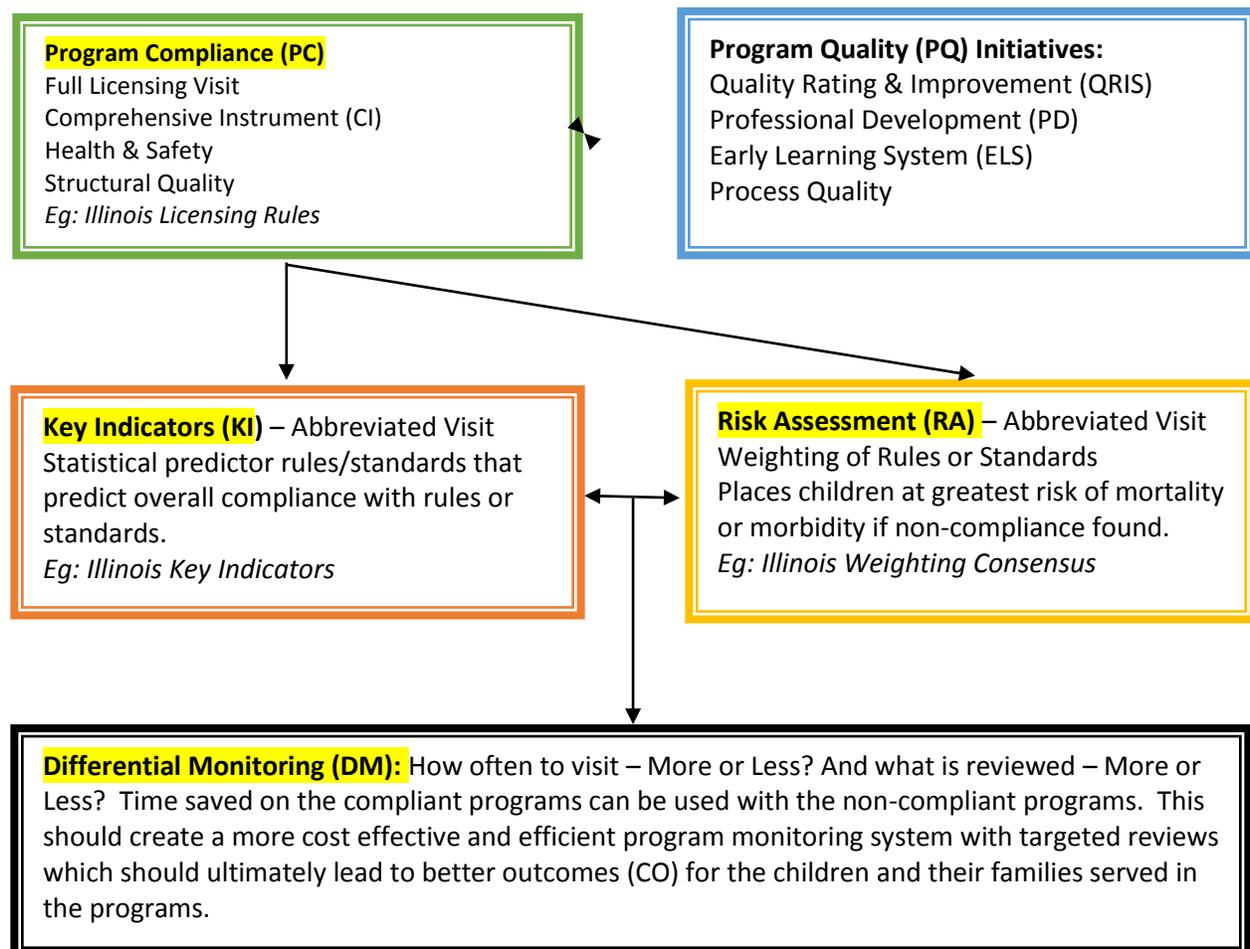
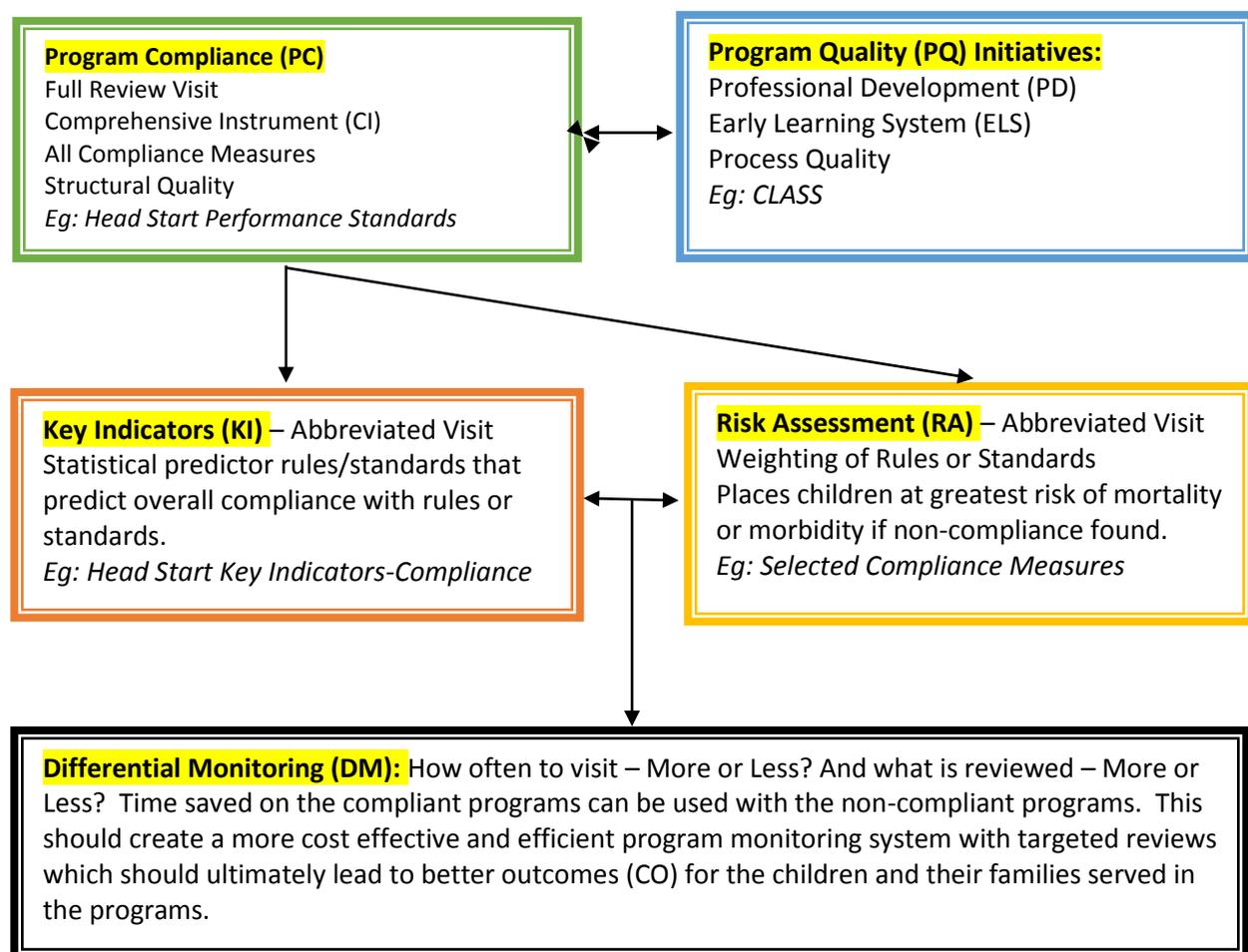


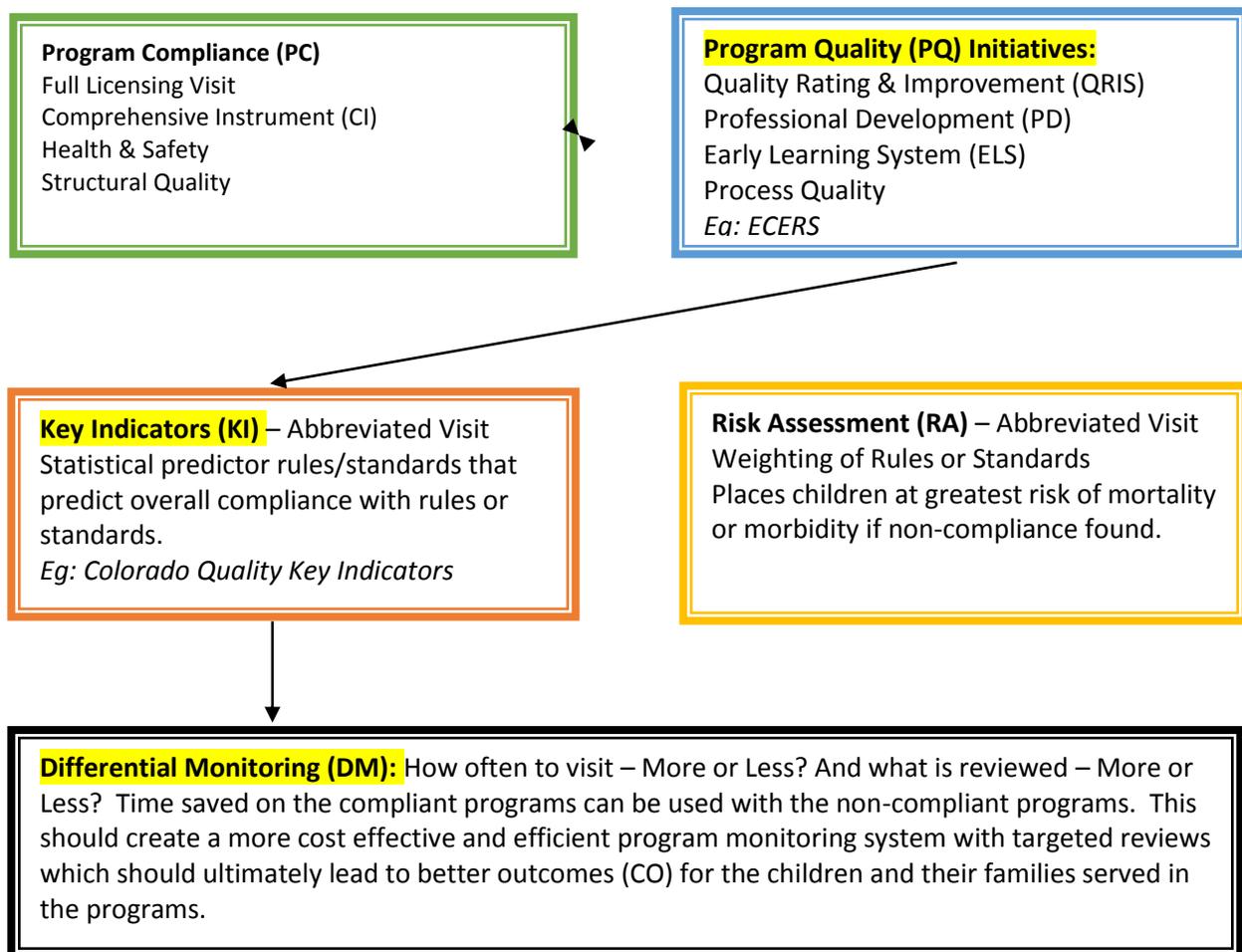
Figure 6 depicts the new aligned differential monitoring system being employed in Head Start. Head Start has a very comprehensive system that employs various aspects from all the components in their system. The Head Start Performance Standards are very comprehensive, CLASS is used as a major process quality measure and both a key indicator (Head Start Key Indicator – Compliance (HSKI-C)) and risk assessment (Selected Compliance Measures) are utilized in their program monitoring system. The Head Start new Aligned Program Monitoring system comes closest to the comprehensive national model.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Head Start (HS)
Figure 6**



In Figure 7 a very different scenario played out in the state of Colorado in which key indicators were developed for their QRIS system rather than for their licensing system. As mentioned earlier, when applying the key indicator methodology to Quality Initiatives one needs to be very cautious if the data distribution is not exceptionally skewed as is the case with licensing data. Some of the data were sufficiently skewed to be able to be used in generating quality key indicators there were limitations noted.

**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©)(Fiene, 2014)
Colorado Example (CO)
Figure 7**



This paper presents the latest examples of national and state agencies differential monitoring approaches. It clearly demonstrates that there are many different approaches to developing and implementing differential monitoring. A key research question for the future as more states utilize the different approaches is to study if one approach is better than the next or a combination works better than most. From my 40+ years of experience as a researcher and state policy analyst I would suggest that a more comprehensive approach that employs the full menu of program quality initiatives similar to the Head Start or the NYQI approaches will be most effective.

As mentioned in the introduction of this paper in describing the Comprehensive National Example of the DMLM© Model the following three tables (Tables 1-3) present a Differential Monitoring Scoring Protocol (DMSP©) that can potentially be used to compare states on how in depth their differential monitoring system is. Table 1 describes the DMSP© in narrative terms delineating the various systems that need to be in place in order to get a particular score. A score of 0 means no systems are in place or do not intersect while a score of 10 means that all of the systems are in place and intersect or are linked. Table 2 gives the points assigned to the specific systems that are part of a differential monitoring system. And Table 3 gives the actual points assigned to the state & national examples that have been presented in this paper for *New York (NY)*, *Georgia (GA)*, *Head Start (HS)*, *Kansas (KS)*, *Illinois (IL)*, and *Colorado (CO)*. The total points assigned to the comprehensive model are also provided as a point of context.

There are a couple of important things to note about the DMSP© in Table 2, such as: if Key Indicators (KI) and Risk Assessment (RA) are linked, it negates KI and RA being scored separately. This is depicted by those cells being blocked out in red. If KI and RA are developed separately, it is very improbable that they will not be linked but that is always a possibility, so it is listed as so. Linking Program Compliance/Licensing (PC) and Program Quality (PQ) Initiatives is a highly desirable event and is assigned a high score (4 points). Linking KI and RA is also considered a highly desirable event and is assigned a high score (4 points).

Notes a, b, c: The arrows going from Key Indicators (KI) and Risk Assessment (RA) to Differential Monitoring (DM) can be configured in the following ways: only KI (Kansas); only RA (don't have an example of this as of this writing) or a combination of KI and RA (Illinois) but this configuration could mean all of the KI and RA rules which would be more rules than if only KI or RA rules were selected or only those rules that overlap (KI+RA) which would be a much reduced number of rules. Or a different configuration determined by the state agency.

Table 1: Differential Monitoring Scoring Protocol (DMSP)© (Fiene, 2014)

Score	Systems Present
0	No systems in place.
2	KI or RA in place and not linked.
4	(KI & RA in place but not linked) or (PC + PQ are linked).
6	(KI & RA in place) & (KI + RA are linked)
8	(KI & RA in place but not linked) & ((PC + PQ) are linked).
10	All systems in place and linked.

Table 2: Differential Monitoring Scoring Protocol (DMSP)© Point Assignment (Fiene, 2014)

Score	Systems Present and Point Assignment
0	No systems in place.
2	(KI (1)) & (KI -> DM (1)) or ((RA (1)) & (RA -> DM (1)))
4	(PC + PQ (4)) or (KI (1) & (KI -> DM (1)) & (RA (1) & (RA -> DM (1)))
6	(KI + RA -> DM (4)) & (KI (1)) & (RA (1))
8	(KI (2) & RA (2)) & (PC + PQ (4)).
10	(KI + RA -> DM (4)) & (KI (1)) & (RA (1)) & (PC + PQ (4))

KI (Key Indicators); RA (Risk Assessment); PC (Program Compliance/Licensing); PQ (Program Quality Initiatives)

Table 3: DMLM© SCORING PROTOCOL WITH STATE EXAMPLES

SYSTEMS (pts)	MODEL	GA	NY	HS	IL	KS	CO
KI (1)	1	-	1	1	1	1	1
RA (1)	1	1	1	1	1	-	-
KI + RA -> DM (4) KI + RA (2)	4	2	4	4	4		
PC + PQ (4)	4	4	-	4	-	-	-
KI -> DM (1)						1	1
RA -> DM (1)		1				-	-
TOTAL (10)	10	8	6	10	6	2	2

GA (Georgia); NY (New York); HS (Head Start); IL (Illinois), KS (Kansas); CO (Colorado)

**DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A
NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL
(ECPQIM⁴©) FOR EARLY CARE AND EDUCATION REGULATORY
AGENCIES**

Richard Fiene, Ph.D.

**Professor of Human Development & Family Studies and Psychology (Retired)
Bennett Pierce Prevention Research Center
The Pennsylvania State University**

**Research Institute for Key Indicators *LLC* (RIKI)
41 Grandview Avenue
Middletown, Pennsylvania USA 17057**

**717-944-5868 Phone and Fax
DrFiene@gmail.com**

Suggestion citation: Fiene (2015). *Differential monitoring logic model (DMLM©): A new early childhood program quality indicator model (ECPQIM⁴©) for early care and education regulatory agencies*, Middletown, PA: Research Institute for Key Indicators (RIKI).

DIFFERENTIAL MONITORING LOGIC MODEL (DMLM©): A NEW EARLY CHILDHOOD PROGRAM QUALITY INDICATOR MODEL (ECPQIM⁴©) FOR EARLY CARE AND EDUCATION REGULATORY AGENCIES

ABSTRACT

A new Early Childhood Program Quality Indicator Model (ECCPQIM⁴©) is described which utilizes targeted program monitoring (Differential Monitoring) via two licensing methodologies: Key Indicators and Risk Assessments. The theoretical and conceptual framework as well as a logic model are presented along with a scoring protocol that can be utilized to compare state/province and national organizations on how they are designing and implementing their program monitoring systems. A state/province/national framework/plan is presented as well as results from five (5) states (Georgia, Kansas, Illinois, Colorado, and New York) and a national organization (Office of Head Start). The five states and national organization are then compared using the Differential Monitoring Scoring Protocol (DMSP©). The Head Start program monitoring system scored a perfect 10 out of 10 in utilizing the DMSP©. Suggestions are made in how the scoring protocol could be used for making comparisons internationally and for future research in comparing various approaches.

Key Words: Program Monitoring, Differential Monitoring, Program Quality, Licensing.

Background

This paper will introduce a Differential Monitoring Logic Model (DMLM©) which provides a new Early Childhood Program Quality Indicator Model (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 (see Figures 1 & 2 for a graphical depiction of the theoretical underpinnings and actual design & logic model for the ECPQIM⁴©/DMLM).

The DMLM© can be used by early care and education state/province agencies, Federal agencies, and large provider organizations where an economy of scale is required. This model can be used with state as well as national standards, such as state licensing rules/regulations and *Caring for Our Children* (AAP, 2012). Most states and Federal agencies have either some or all of the key elements of this model in their overall monitoring systems. The purpose of this model is to alter a one-size fits all monitoring system to one that is targeted, spending more time with problem programs who need additional assistance. This is a cost neutral model that is both cost effective and efficient and re-allocates resources from the compliant programs to the non-compliant programs. Presently there is not a measurement rubric for making comparisons within the USA or internationally when it comes to measuring the effectiveness and efficiency of child care and

early care program monitoring systems. This can become a very important tool as the USA begins implementation of the re-authorization of the Child Care and Development Block Grant.

Insert Figure 1

The ECPQIM⁴©/DMLM© is based very heavily in translational research and implementation science as a means of building an ongoing program monitoring system based upon the latest empirical demonstrations in the early care and education research literature. It is at the intersection of child care public policy, early care and education interventions, and empirical research. The ECPQIM⁴©/DMLM© along with the scoring protocol introduced in this paper could provide a framework for making comparisons amongst states/provinces, national organizations, and countries in how they have designed and implemented their respective program monitoring of child care and early care & education systems similar to how Child Care Aware has developed a reporting format for the USA in comparing states on regulatory and oversight functions. The author reported on such a comparison in a previous study in an earlier edition of this journal (Fiene, 2013). The DMLM© framework and scoring protocol could provide a similar measurement tool for assessing child care and early childhood education program monitoring systems.

DMLM© Key Elements (see Figure 2): **CI** = state or federal child care standards, usually rules or regulations that measure health and safety - *Caring for Our Children* (AAP, 2012) will be applicable here. **PQ** = Quality Rating and Improvement Systems (QRIS) standards at the state level; process quality measures. **RA** = risk assessment tools/systems in which only the most critical rules/standards are measured. *Stepping Stones* (NRC, 2013) 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, 2002) 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.

Insert Figure 2

Once the above key elements are in place, it is then possible to look at the relationships (this is depicted by the arrows that go from one box to another) amongst them to determine if the system is operating as it was intended; in other words, to determine if the DM system is improving the health, safety, program quality and ultimately the overall development of the children it serves.

In the Methodology section, a scoring protocol (DMSP© - Differential Monitoring Scoring Protocol©) is introduced which attempts to quantify these relationships and to give us a means for making measurements and comparisons across various types of organizations.

The DMLM© provides a cross-cutting methodology that can be used in all child care/early care and education delivery systems as well as in other human services. In the past many of these monitoring systems have functioned in silos. The DMLM© integrates all these various monitoring systems together so that the overall monitoring system can be validated as being cost effective and efficient. This can be an important development as available funds become more scarce in the future as international organizations deal with fewer and fewer resources.

Methods

National/State/Provincial Agency Plan for implementing a Differential Monitoring System:

The **first step** in utilizing the DMLM© for a state/province/nation 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/province/nation. If the state/province/nation does not utilize a tool or checklist or does not review all violation data than it needs to consider these changes because the DMLM© is based upon an Instrument Based Program Monitoring System (IPM)(Fiene & Nixon,1985) which utilizes tools/checklists to collect data on all rules.

The **second step** for the state/province/nation is to compare their nation's/state's/province's rules

with the National *Health and Safety Performance Standards (Caring for Our Children)*(AAP, 2012) or an equivalent international set of standards to determine the overlap and coverage between the two.

The **third step** for the state/province/nation if it utilizes a Risk Assessment (RA) tool is to assess the relationship between this tool and *Stepping Stones* (NRC, 2013) or an equivalent international set of targeted standards to determine the overlap and coverage between the two.

The **fourth step** for the state/province/nation is to compare the results from the CI with the RA tools.

In the **fifth step**, if a state/province/nation 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 that measure process quality. This is a very important step because very few empirical demonstrations appear in the research literature regarding this relationship.

The **sixth step** is for the state/province/nation 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. If a state/province/nation 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/nation 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. Also, this step should drive decisions within the technical assistance/training/professional development (PD) system in what resources are allocated to a particular facility.

The **eighth and final step** for the state/nation 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/provinces/nations at this point have this capability on a large scale. However, as Early Learning Networks/Systems and Standards (ELS) are developed, this will become more common place.

The ECPQIM⁴©DMLM© is presented without two additional items that were present in the 2012/2013 versions which are important to note. The algorithm (Fiene, 2012, 1013) and validation framework (Zellman & Fiene, 2012) are not presented because the author felt that these two components took away from a more direct presentation of differential monitoring. For those interested readers, please refer to my previous abstracts (Fiene, 2012, 2013) which

included the algorithm and validation frameworks.

Just another brief word about the Theoretical Underpinnings for ECPQIM⁴. This graphic (Figure 1) attempts to provide the relationships amongst public policy, interventions, and empirical evidence through the lens of translational research, implementation science, and program monitoring. In constructing the ECPQIM⁴ concepts were borrowed from each area and integrated them in a model for monitoring early care and education programs. The graphic provides a means for displaying the relationships and potential intersections as well as the content that is important to each scientific/research field.

Figure 3 is provided as additional information regarding differential monitoring conceptually without all the details as in figure 2; and figure 4 is provided to demonstrate the impact that a state's/provincial/national licensing law can have on using the Key Indicators and Risk Assessment methodologies.

Insert Figures 3 & 4

Also, taking Figure 2 and attempting to quantify these relationships, a scoring protocol is proposed as depicted in Table 1. This can provide a numerical means of comparing various

differential monitoring systems and their relative comprehensiveness. This protocol could be a useful tool in future research for determining which combinations work best.

Insert Table 1

The next section provides the results from a national organization and five states who used the above methodology to implement their respective differential monitoring systems.

Results and Discussion

The Early Childhood Program Quality Indicator Model (ECPQIM©) and its latest iteration presented as a logic model: Differential Monitoring Logic Model (DMLM©) have been written about extensively by this author (Fiene & Nixon, 1985; Griffin & Fiene, 1996; Fiene & Kroh, 2000; Fiene, 2013). Several states and Head Start have used the model in order to re-align their program monitoring systems. This paper presents the results of those new program monitoring systems through the lenses of the ECPQIM©/DMLM© logic model display. Each particular approach used various components of the overall comprehensive national model and have been highlighted by connecting arrows. It is proposed that this approach could be applied at an international level as well.

The interested reader should obtain a copy of the Office of Child Care's *Licensing Brief on Differential Monitoring, Risk Assessment, and Key Indicators* published by the National Center on Child Care Quality Improvements which gives additional details regarding these approaches and methodologies as well as other state examples. Please go to the following URL website: (https://childcareta.acf.hhs.gov/sites/default/files/1408_differential_monitoring_final_1.pdf). In fact, this paper builds upon that excellent *Licensing Brief*.

Let's start with Figure 5 which provides the Comprehensive National Example that depicts all the possible interconnections and gives national examples from the research literature. As one will see, it is possible for a national organization or a state/provincial agency to select the various components from the model based upon what is available in their particular organization. All do have the program compliance/licensing component (PC) but not all have fully functional program quality initiatives (PQ) or do not have the data to draw from the program quality initiatives.

The next level of components are the key indicator (KI) and risk assessment (RA) approaches or methodologies which organizations or state agencies can use alone or in tandem. One limitation in the key indicator methodology is not to use it with program initiatives if the data are not severely skewed in their data distribution as is the case with licensing data.

The last component is the resulting differential monitoring (DM) approach based upon the results

from using the key indicator and risk assessment methodologies either alone or in tandem. This is the ultimate revision of the program monitoring system in which how often and what is reviewed are answered.

All the components are highlighted (this is indicated by the arrows going from one box to another) in Figure 5 because all are possibilities to be used by a national or state agency. The examples in Figure 5 are drawn from the national research literature so *Caring for Our Children* (AAP, 2012) is the example for Program Compliance, Licensing, and the Health & Safety Comprehensive Instrument (CI). The following examples in Figures 6-11 will show some differences in how national and state agencies have developed their respective differential monitoring systems through their use of key indicator (KI) and risk assessment (RA) methodologies, and linking their licensing/program compliance (PC) and program quality (PQ) initiatives. Tables 1-3 explain the scoring protocol and provide results from the national Head Start program and five states geographically dispersed around the USA (New York, Georgia, Illinois, Kansas, and Colorado). Also see the end of the paper for an explanation of Notes a,b,c in Figure 5.

Insert Figure 5

Figure 6 provides an example from New York (NY) where the state agency is attempting to restructure their early care and education program monitoring system to have a better balance between licensing and key program quality indicators. The plan is to have licensing staff collect data from both areas which means a need to save time in the licensing reviews via key indicators and to only identify indicators of quality through a risk assessment approach. The results from these two methodologies will then be combined into a Quality Indicators Instrument to be used by licensing staff in their annual reviews.

Insert Figure 6

Figure 7 provides an example from Georgia (GA) in which the driving methodology is a risk assessment core rule review system that results in a differential monitoring system called the Annual Compliance Determination Worksheet (ACDW) approach. Key indicators are not used directly but were used as part of the risk assessment core rule development. Please note how the relationship amongst the various components is different from the NY approach delineated in Figure 6. There is a link to their program quality initiatives which proved very significant in the validation studies performed on their Core Rule differential monitoring system.

Insert Figure 7

Figure 8 presents a very different approach from the previous two approaches. In Kansas's (KS) case, the state agency was only interested in developing a key indicator approach and was not interested in risk assessment nor had the capability to tie data together from their program quality initiatives. This is noted by the arrow connections which is more minimal in this depiction. As one can see, this still is a viable option for developing a differential monitoring approach.

Insert Figure 8

Figure 9 depicts the use of both key indicator and risk assessment methodologies in Illinois (IL) with their licensing system but no data interaction with their program quality initiatives. It is proposed that both methodologies will be used together in future licensing reviews of programs which will constitute their differential monitoring system approach.

Insert Figure 9

Figure 10 depicts the new aligned differential monitoring system being employed in Head Start (HS). Head Start has a very comprehensive system that employs various aspects from all the components in their system. The Head Start Performance Standards are very comprehensive, CLASS is used as a major process quality measure and both a key indicator (Head Start Key Indicator – Compliance (HSKI-C)) and risk assessment (Selected Compliance Measures) are utilized in their program monitoring system. The Head Start new Aligned Program Monitoring system comes closest to the comprehensive national model.

Insert Figure 10

In Figure 11 a very different scenario played out in the state of Colorado (CO) in which key indicators were developed for their QRIS system rather than for their licensing system. As mentioned earlier, when applying the key indicator methodology to Quality Initiatives one needs to be very cautious if the data distribution is not exceptionally skewed as is the case with licensing data. Some of the data were sufficiently skewed to be able to be used in generating

quality key indicators but there were limitations noted.

Insert Figure 11

The above results clearly demonstrate how agencies can take very different approaches to designing and implementing their differential monitoring system. The next research question is to determine if agencies that have higher scores (more than 6) if they are more effective and efficient than those agencies that have lower scores (less than 5).

Conclusion

This paper presents the latest examples of national and state agencies differential monitoring approaches. It clearly demonstrates that there are many different approaches to developing and implementing differential monitoring. A key research question for the future as more states utilize the different approaches is to study if one approach is better than the next or a combination works better than most. From 40+ years of experience as a researcher and state policy analyst I would suggest that a more comprehensive approach which employs the full menu of program quality initiatives similar to the Head Start or the New York approaches will be most effective.

As mentioned in the introduction of this paper in describing the Comprehensive National Example of the DMLM© Model Tables 1-3 present a Differential Monitoring Scoring Protocol (DMSP©) that can potentially be used to compare states on how in depth their differential monitoring system is. Table 1 describes the DMSP© in narrative terms delineating the various systems that need to be in place in order to get a particular score. A score of 0 means no systems are in place or do not intersect while a score of 10 means that all of the systems are in place and intersect or are linked. Table 2 gives the points assigned to the specific systems that are part of a differential monitoring system. And Table 3/Figure 12 give the actual points assigned to the state & national examples that have been presented in this paper for *New York (NY)*, *Georgia (GA)*, *Head Start (HS)*, *Kansas (KS)*, *Illinois (IL)*, and *Colorado (CO)*. The total points assigned to the comprehensive model are also provided as a point of context.

There are a couple of important things to note about the DMSP© in Table 2, such as: if Key Indicators (KI) and Risk Assessment (RA) are linked, it negates KI and RA being scored separately. If KI and RA are developed separately, it is very improbable that they will not be linked but that is always a possibility, so it is listed as so. Linking Program Compliance/Licensing (PC) and Program Quality (PQ) Initiatives is a highly desirable event and is assigned a high score (4 points). Linking KI and RA is also considered a highly desirable event and is assigned a high score (4 points).

Insert Tables 2 & 3 and Figure 12

For future research, it will be interesting to see if this ECPQIM⁴©/DMLM© model has applicability from an international perspective. Some of the key elements present in USA state systems are organized very differently in other countries and would have to be adjusted. Also, it will be interesting to see if the DMSP© can be developed as a scoring systems similar to the Child Care Aware Report Card Benchmarks protocol where it will be possible to make comparisons across state and national agencies.

Endnotes a, b, c:

The arrows going from Key Indicators (KI) and Risk Assessment (RA) to Differential Monitoring (DM) can be configured in the following ways: only KI (Kansas); only RA (don't have an example of this as of this writing) or a combination of KI and RA (Illinois) but this configuration could mean all of the KI and RA rules which would be more rules than if only KI or RA rules were selected or only those rules that overlap (KI+RA) which would be a much reduced number of rules. Or a different configuration determined by the state agency.

SENDING00: ECPQIM – DMLM – ICEP1dI (2)aC RIKI HF

References

AAP (2012). *National health and safety performance standards*, Evanston, IL: American Academy of Pediatrics.

Fiene (2013). A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks, *International Journal of Child Care and Education Policy*, 7(1), 1-15.

Fiene (2012/2013). *Early childhood program quality indicator model (ECPQIM⁴©): Differential monitoring logic model and algorithm (DMLMA©)*, Middletown, PA: Research Institute for Key Indicators LLC (RIKI).

Fiene (2002). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.

Fiene & Kroh (2000). *Licensing Measurement and Systems, NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.

Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, 14(3), 198-214.

Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.

NRC (2013). *Stepping stones to caring for our children*. Denver, Colorado: National Resource Center for Health and Safety in Child Care.

Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Figure 1

The Theoretical Underpinnings for ECPQIM⁴: Early Childhood Program Quality Indicator Model©

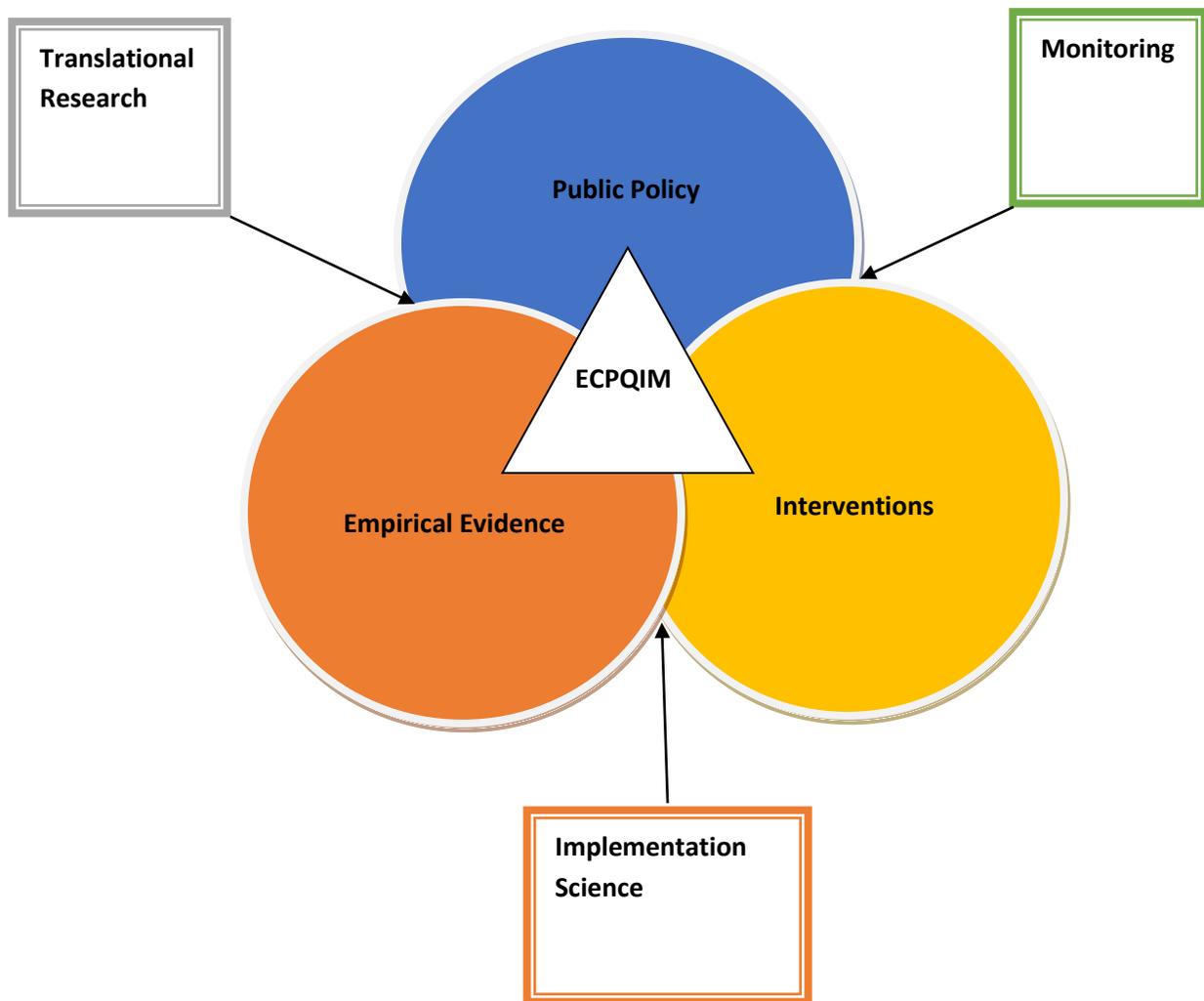


Figure 2

**Early Childhood Program Quality Indicator Model (ECPQIM⁴©):
Differential Monitoring Logic Model (DMLM©)
Comprehensive National Example**

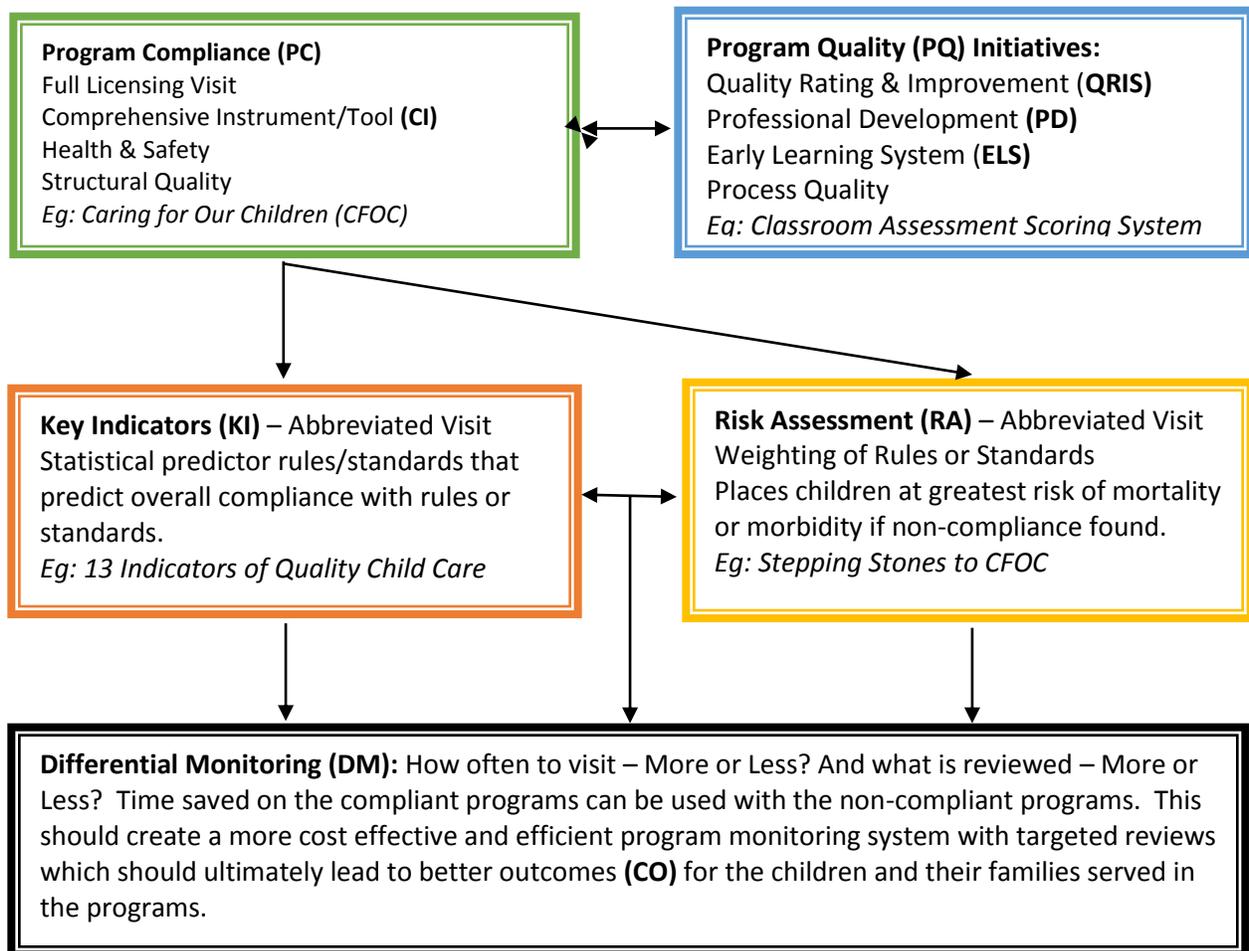


Figure 3

Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators

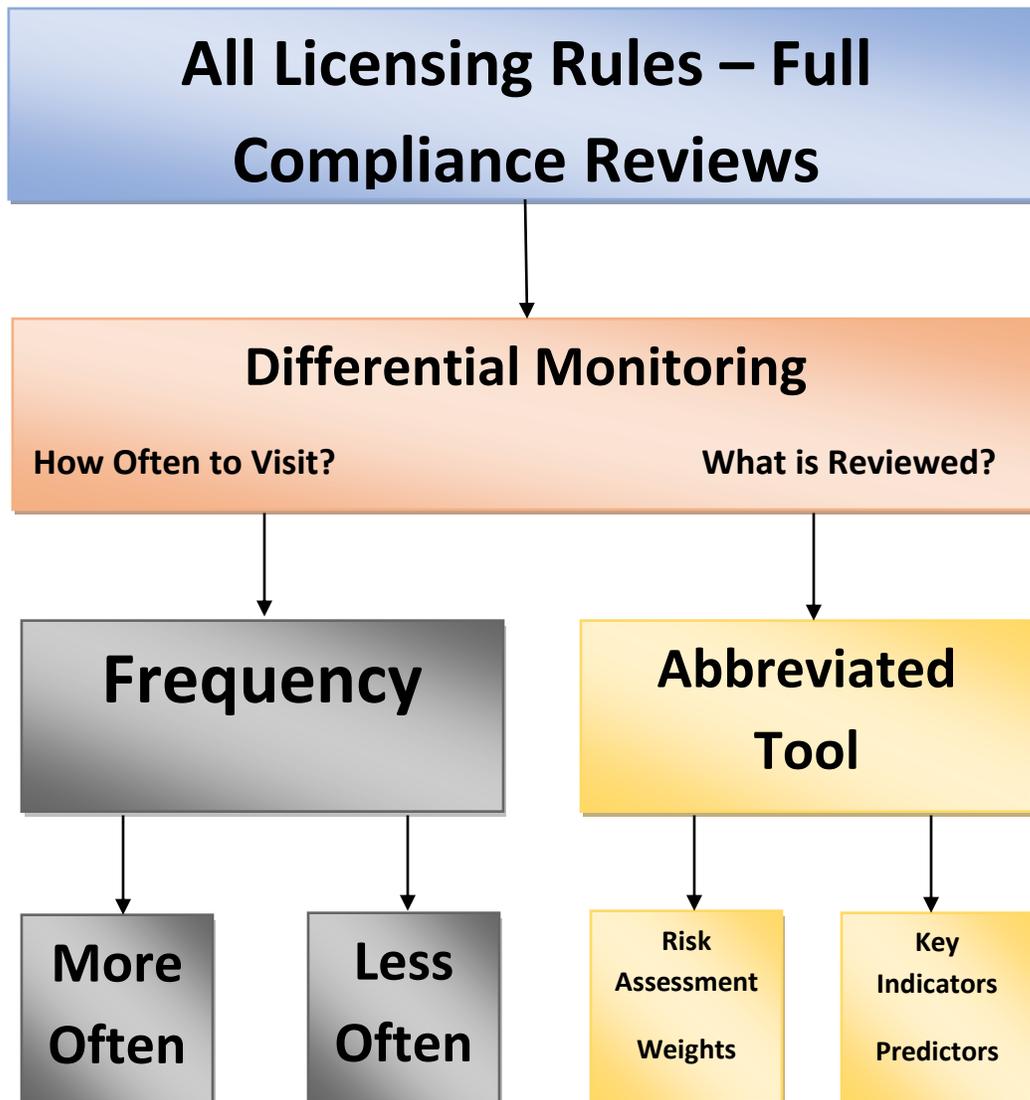
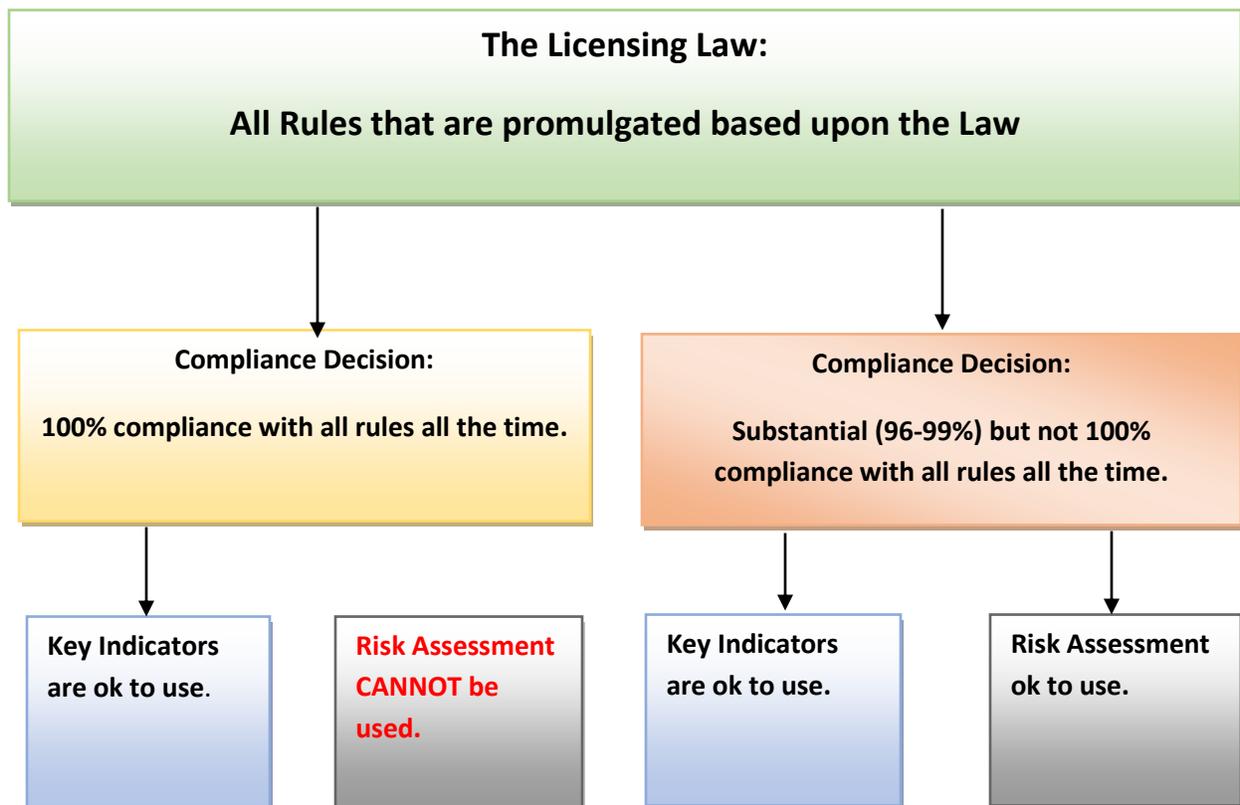


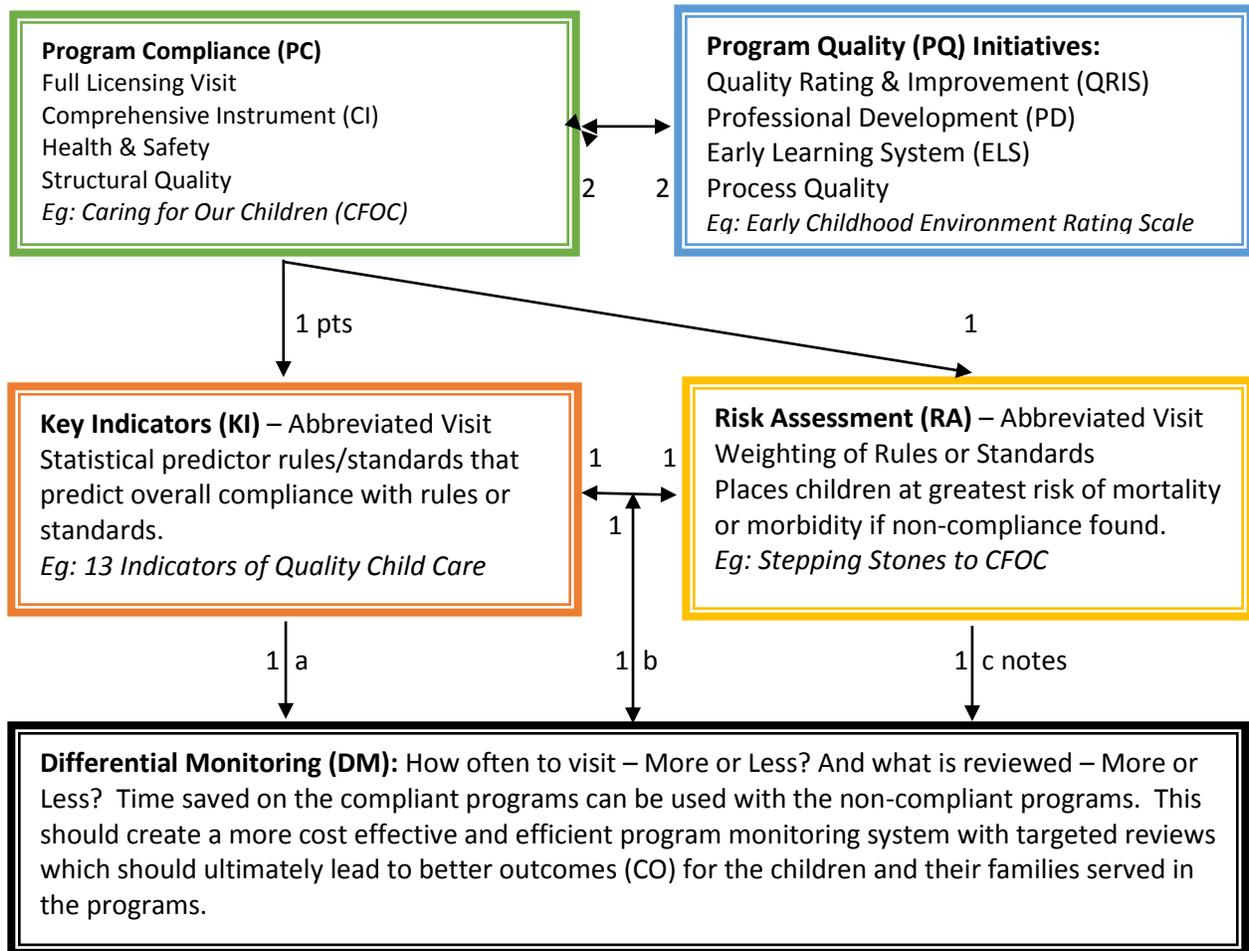
Figure 4

When Key Indicators and Risk Assessments Can Be Used

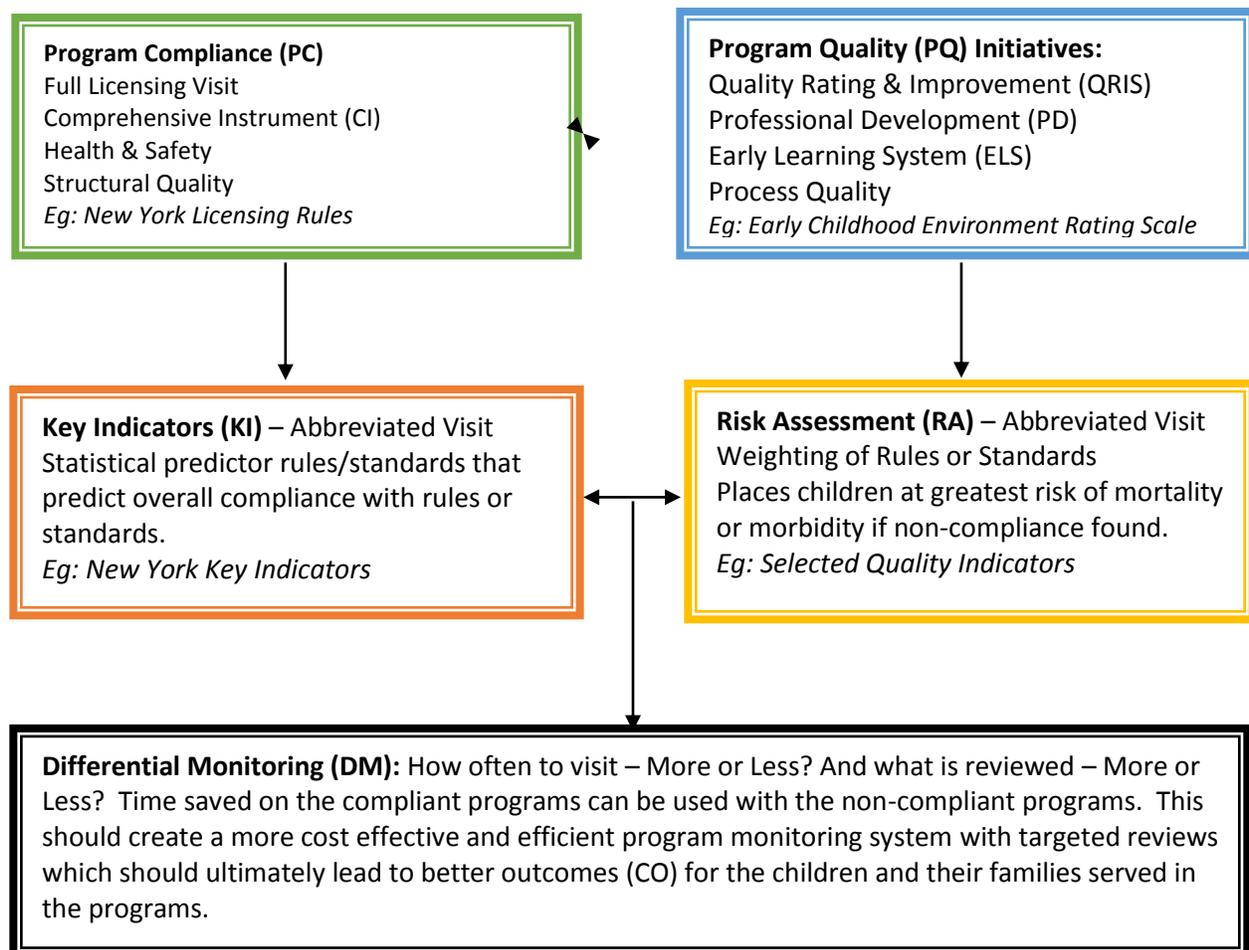


**Early Childhood Program Quality Indicator Model (ECPQIM4©):
Differential Monitoring Logic Model (DMLM©) Comprehensive National
Scoring Protocol Example (Maximum of 10 Points)**

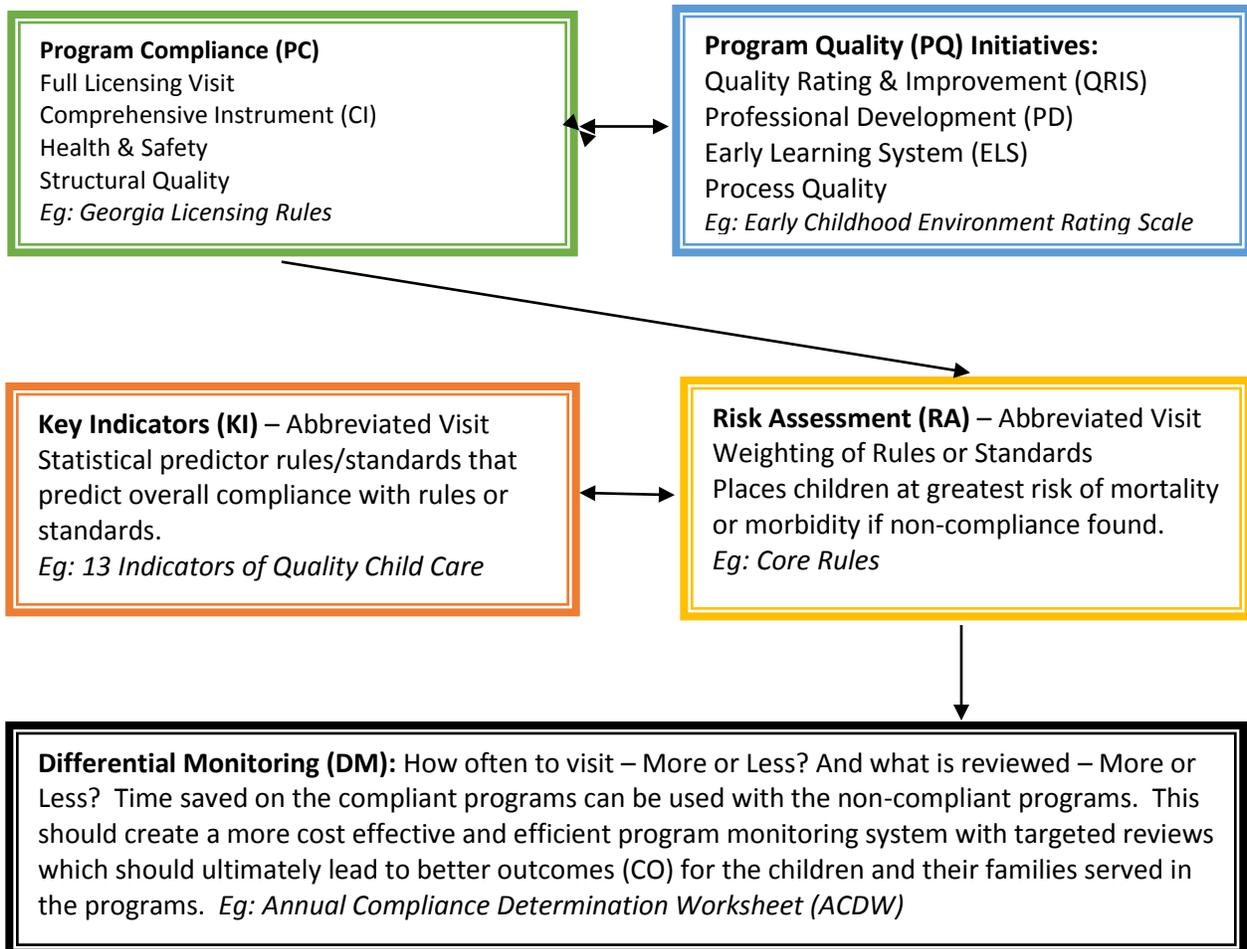
Figure 5



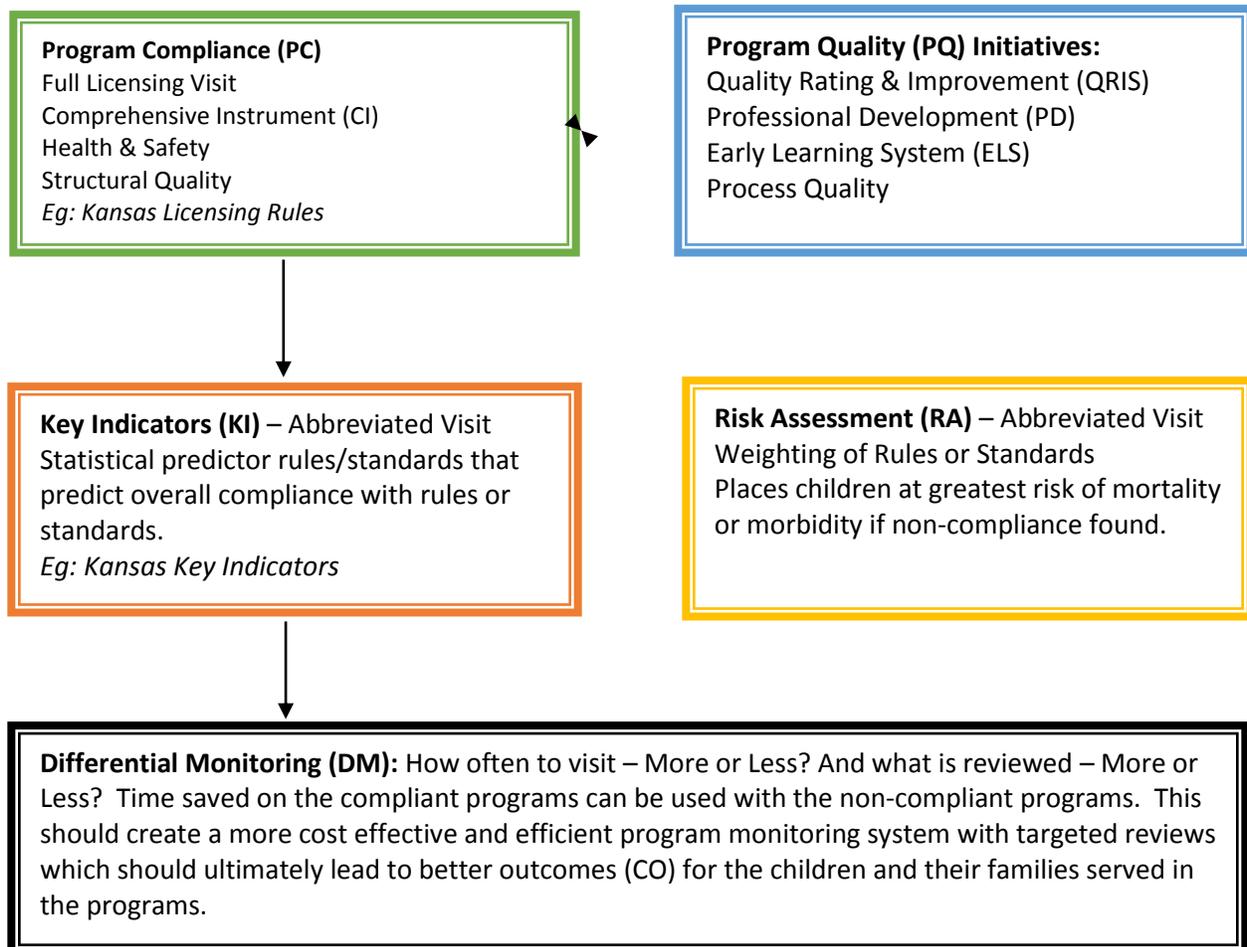
(ECPQIM4©)(DMLM©): New York Example (NY)
Figure 6



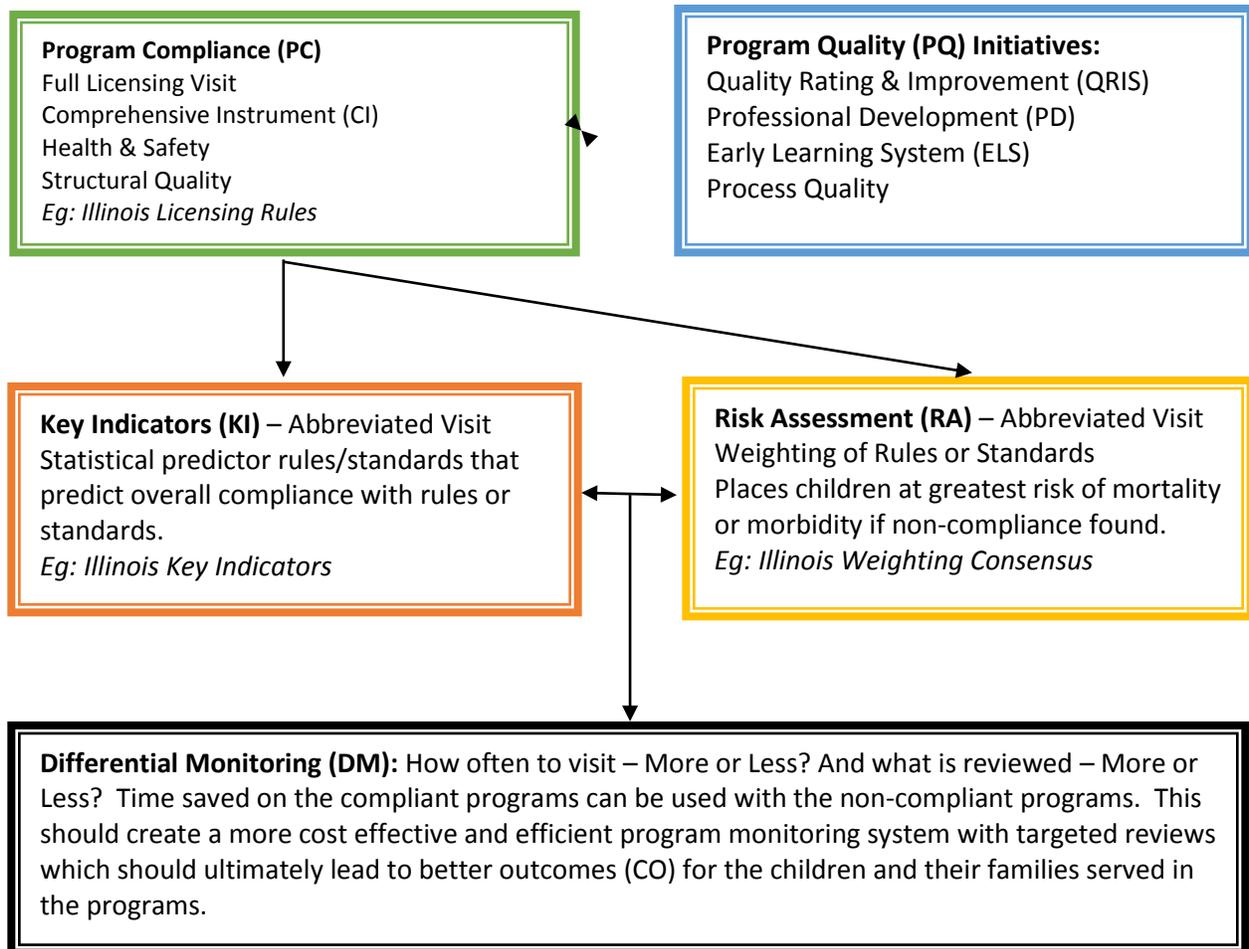
(ECPQIM4©)(DMLM©): Georgia Example (GA)
Figure 7



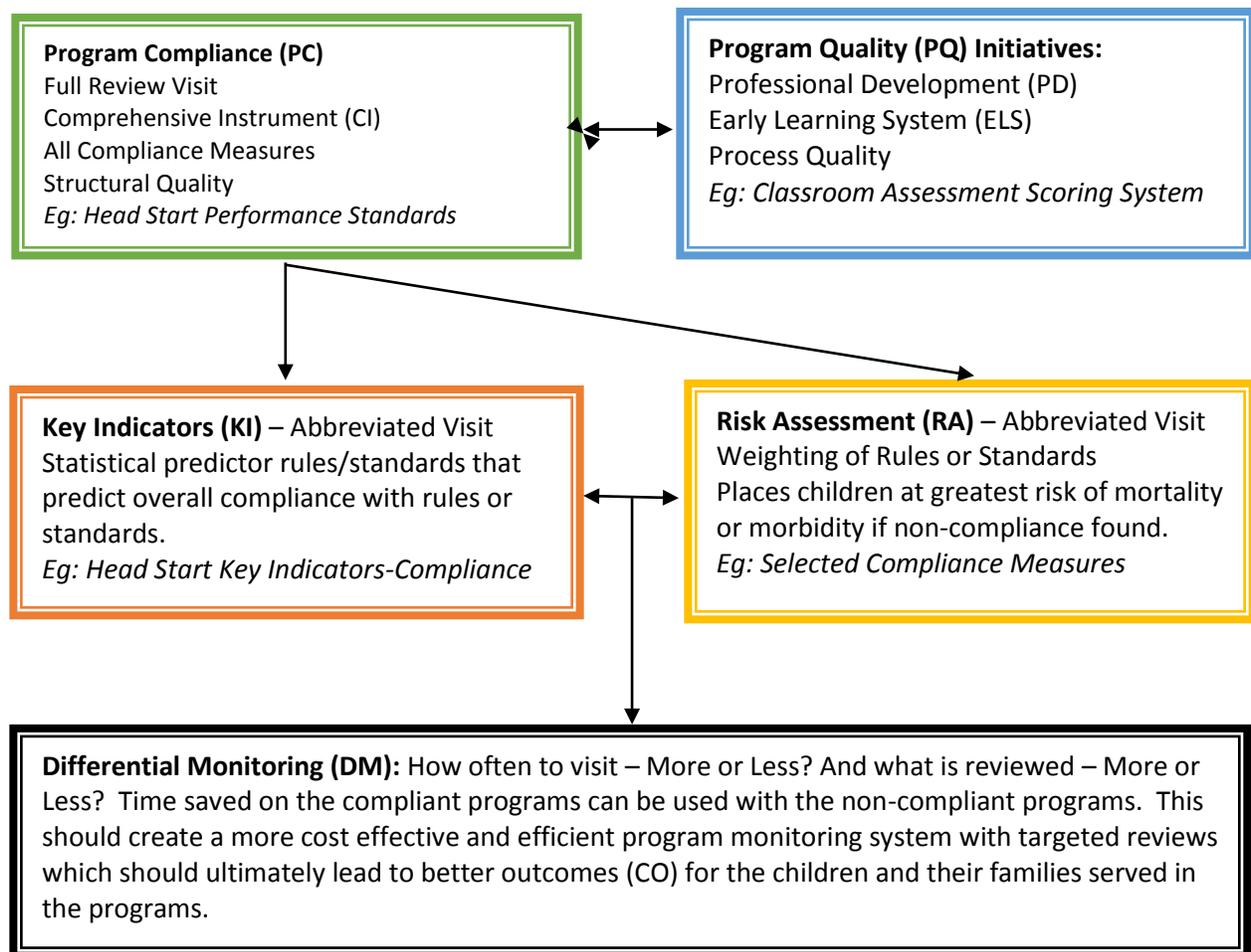
(ECPQIM4©)(DMLM©): Kansas Example (KS)
Figure 8



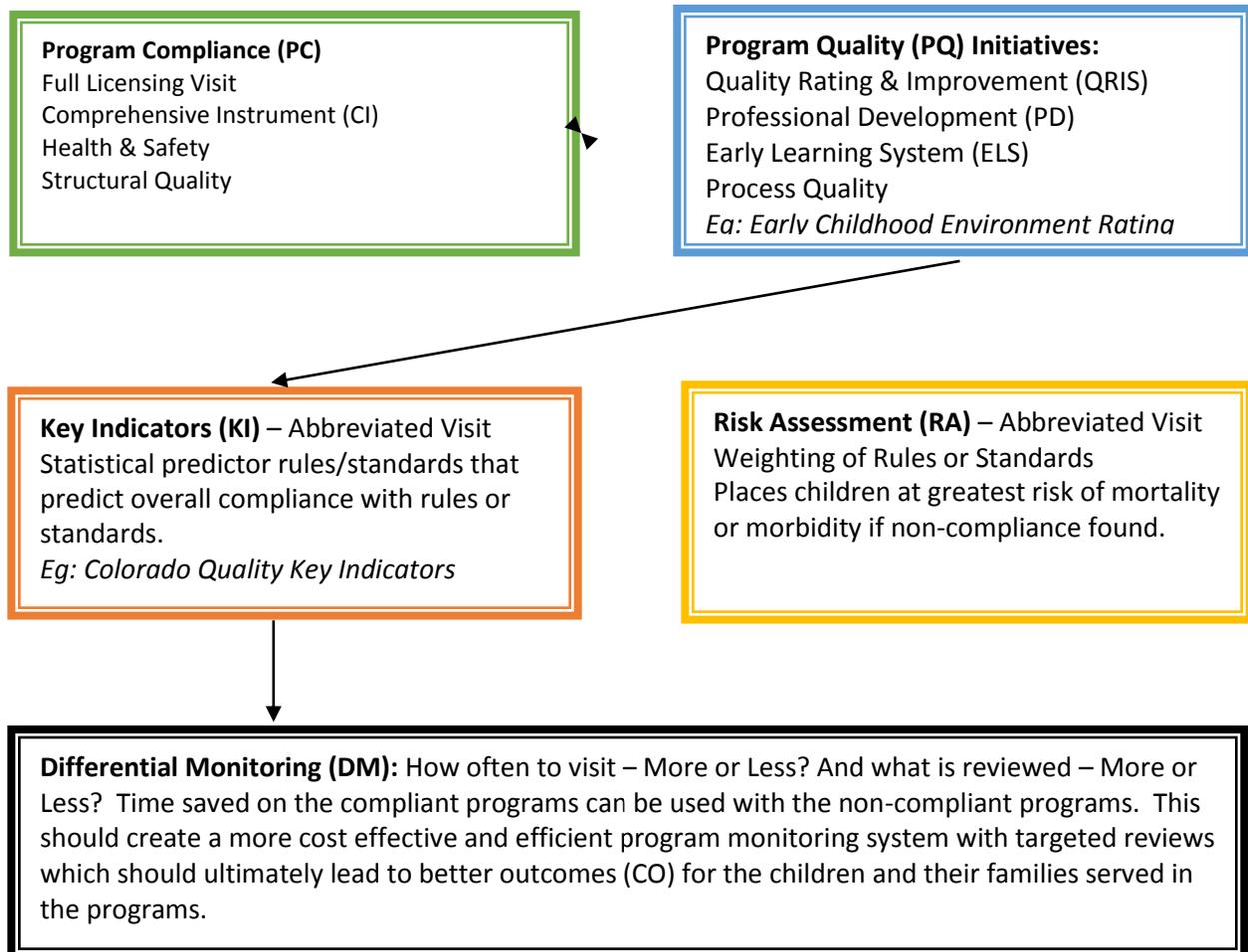
(ECPQIM4©)(DMLM©): Illinois Example (IL)
Figure 9



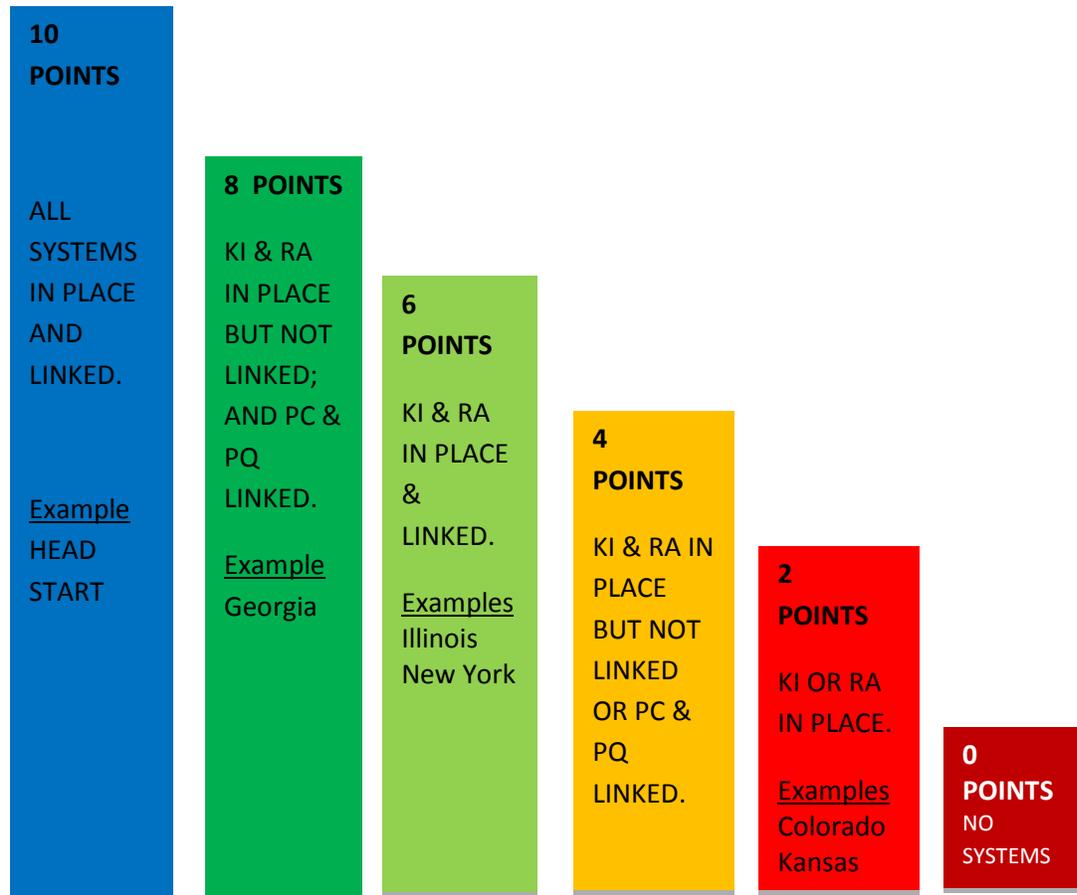
(ECPQIM4©)(DMLM©): Head Start Example (HS)
Figure 10



**(ECPQIM4©)(DMLM©): Colorado Example (CO)
Figure 11**



DMSP© SCORING PROTOCOL WITH STATE AND NATIONAL AGENCIES AS EXAMPLES
Figure 12



SCORING

KI = Key Indicators; RA = Risk Assessment; PC = Licensing; PQ = Program Quality Initiatives

Table 1: Differential Monitoring Scoring Protocol (DMSP)©

<i>Score</i>	<i>Systems Present</i>
0	No systems in place.
2	KI or RA in place and not linked.
4	(KI & RA in place but not linked) or (PC + PQ are linked).
6	(KI & RA in place) & (KI + RA are linked)
8	(KI & RA in place but not linked) & ((PC + PQ) are linked).
10	All systems in place and linked.

KI (Key Indicators); RA (Risk Assessment); PC (Program Compliance/Licensing); PQ (Program Quality Initiatives)

Table 2: Differential Monitoring Scoring Protocol (DMSP)© Point Assignment

<i>Score</i>	<i>Systems Present and Point Assignment</i>
0	No systems in place.
2	(KI (1)) & (KI -> DM (1)) or ((RA (1)) & (RA -> DM (1)))
4	(PC + PQ (4)) or (KI (1) & (KI -> DM (1)) & (RA (1)) & (RA -> DM (1)))
6	(KI + RA -> DM (4)) & (KI (1)) & (RA (1))
8	(KI (2) & RA (2)) & (PC + PQ (4)).
10	(KI + RA -> DM (4)) & (KI (1)) & (RA (1)) & (PC + PQ (4))

KI (Key Indicators); RA (Risk Assessment); PC (Program Compliance/Licensing); PQ (Program Quality Initiatives)

Table 3: DMLM© SCORING PROTOCOL WITH STATE EXAMPLES

SYSTEMS (pts)	MODEL	GA	NY	HS	IL	KS	CO
<i>KI (1)</i>	1	-	1	1	1	1	1
<i>RA (1)</i>	1	1	1	1	1	-	-
<i>KI + RA -> DM (4)</i> <i>KI + RA (2)</i>	4	2	4	4	4	-	-
<i>PC + PQ (4)</i>	4	4	-	4	-	-	-
<i>KI -> DM (1)</i>	-	-	-	-	-	1	1
<i>RA -> DM (1)</i>	-	1	-	-	-	-	-
TOTAL (10)	10	8	6	10	6	2	2

GA (Georgia); NY (New York); HS (Head Start); IL (Illinois), KS (Kansas); CO (Colorado)

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



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:

Zellman, G. L. & Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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.

Why QRIS validation is important. A QRIS is a primary strategy states employ to improve early childhood education and school-age care (ECE-SAC) program quality. Because ratings are a central element of a QRIS, it is important to collect data to establish that these ratings are accurate and meaningful indicators of quality. Validation studies can lend credibility to a QRIS, identify needed changes, and support continuous improvement of a QRIS.

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; Tenopir, 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

Approach	Activities and Purpose	Typical Questions Approach Addresses	Issues and Limitations
<p>1. Examine the validity of key underlying concepts</p>	<p>Assess whether basic QRIS quality components and standards are the “right” ones by examining levels of empirical and expert support.</p>	<p>Do the quality components capture the key elements of quality?</p> <p>Is there sufficient empirical and expert support for including each standard?</p>	<p>Different QRISs may use different decision rules about what standards to include in the system.</p>
<p>2. Examine the measurement strategy and the psychometric properties of the measures used to assess quality</p>	<p>Examine whether the process used to document and verify each indicator is yielding accurate results.</p> <p>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.</p> <p>Examine the relationships among the component measures to assess whether they are functioning as expected.</p> <p>Examine cut scores and combining rules to determine the most appropriate ways to combine measures of quality standards into summary ratings.</p>	<p>What is the reliability and accuracy of indicators assessed through program administrator self-report or by document review?</p> <p>What is the reliability and accuracy of indicators assessed through observation?</p> <p>Do quality measures perform as expected? (e.g., do subscales emerge as intended by the authors of the measures?)</p> <p>Do measures of similar standards relate more closely to each other than to other measures?</p> <p>Do measures relate to each other in ways consistent with theory?</p> <p>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?</p>	<p>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.¹</p>

Approach	Activities and Purpose	Typical Questions Approach Addresses	Issues and Limitations
<p>3. Assess the outputs of the rating process</p>	<p>Examine variation and patterns of program-level ratings within and across program types to ensure that the ratings are functioning as intended.</p> <p>Examine relationship of program-level ratings to other quality indicators to determine if ratings are assessing quality in expected ways.</p> <p>Examine alternate cut points and rules to determine how well the ratings distinguish different levels of quality.</p>	<p>Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?</p> <p>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?</p>	<p>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.</p>
<p>4. Examine how ratings are associated with children's outcomes.</p>	<p>Examine the relationship between program-level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes.</p>	<p>Do children who attend higher-rated programs have greater gains in skills than children who attend lower-quality programs?</p>	<p>Appropriate demographic and program level control variables must be included in analyses to account for selection factors.</p> <p>Studies could be done on child and program samples to save resources.</p> <p>Findings do not permit attribution of causality about QRIS participation but inferences can be made about how quality influences children's outcomes.</p>

Table 2. Data Needs, Data Sources and Analysis Methods for Selected Studies

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

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

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

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

http://www.acf.hhs.gov/programs/opre/cc/childcare_technical/index.html

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:

Quality Rating Systems: A Key Topic Resource List. New York: Child Care & Early Education Research Connections.

<http://www.researchconnections.org/files/childcare/keytopcis/QualityRatingSystems.pdf>

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

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G. & Boller, K. (2010). **The Child Care Quality Rating System (QRS) Assessment: Compendium of Quality Rating Systems and Evaluations, OPRE Report.** Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

http://www.acf.hhs.gov/programs/opre/cc/childcare_quality/compendium_qrs/qrs_compendium_final.pdf

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.

Lugo-Gil, J., Sattar, S., Boss, C., Boller, K. Tout, K., & Kirby, G. (2011). **The Quality Rating and Improvement System (QRIS) Evaluation Toolkit. OPRE Report #2011-31.** Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

http://www.acf.hhs.gov/programs/opre/cc/childcare_quality/qrisc_toolkit/qrisc_toolkit.pdf

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://qriscnetwork.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

Lahti, M., Langill, C., Sabol, T., Starr, R., & Tout, K., (in progress). **Validating Standards in Child Care Quality Rating and Improvement Systems: Exploring Validation Activities in Four States, OPRE Report.** Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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.

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). ***Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition***. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

http://www.acf.hhs.gov/programs/opre/cc/childcare_technical/reports/complete_compendium_full.pdf

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

American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Academy of Pediatrics, American Public Health Association, & National Resource Center for Health and Safety in Child Care (AAP/APHA/NRCHSCC) (2011). *Caring for our children: National health and safety performance standards guidelines for early care and education programs*. Elk Grove Village, Illinois: American Academy of Pediatrics.

Anastasi, A. (1986). *Psychological testing (5th ed.)*. NY: Macmillan.

Barnard, W., Smith, W., Fiene, R., & Swanson, K. (2006). *Evaluation of Pennsylvania's Keystone STARS quality rating system in child care settings*. Pittsburgh, Pennsylvania: University of Pittsburgh Office of Child Development.

Cizek, Gregory J. Introduction to Validity. Presentation to the National Assessment Governing Board of NAEP. August, 2007.

Elicker, J., Langill, C., Ruprecht, K., & Kwon, K. (2007). *Paths to quality: A child care quality rating system for Indiana: What is its scientific base*. West Lafayette, IN: Purdue University.

Elicker, J., & Thornburg, K. (2011). *Evaluation of quality rating and improvement systems in early childhood programs and school age care: measuring children's development, research to policy, research to practice brief OPRE 2011-11c*. Washington, DC: Department of Health and Human Services, Administration of Children and Families, Office of Planning, Research, and Evaluation.

Elicker, J., Langill, C.C., Ruprecht, K., Lewsader, J., & Anderson, T. (2011). Evaluation of "Paths to QUALITY", Indiana's child care quality rating and improvement system. West Lafayette, IN: Purdue University.

Embretson, S.E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93(1), 179-197.

Fiene, R., Greenberg, M., Bergsten, M., Fegley, C., Carl, B., & Gibbons, E. (2002). *The Pennsylvania early childhood quality settings study*. Harrisburg, Pennsylvania: Governor's Task Force on Early Childhood.

Guion, R.M. (1977). Content validity – The source of my discontent. *Applied Psychological Measurement*, 1, 1-10.

- Halle, T., Vick-Whittaker, J. E., & Anderson, R. (2010). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Harms, T., & Clifford, R. (1989). *Family Day Care Environmental Rating Scale*. New York: Columbia University Teachers College Press.
- Harms, T., Clifford, R., & Cryer, D. (2005). *Early Childhood Environmental Rating Scale – Revised*. New York: Columbia University Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. (2006). *Infant Toddler Environmental Rating Scale – Revised*. New York: Columbia University Teachers College Press.
- Harms, T., Cryer, D. & Clifford, R. (2007). *Family Day Care Environmental Rating Scale – Revised*. New York: Columbia University Teachers College Press.
- Isner, T., Soli, M., Rothenberg, L., Moodie, S., & Tout, K. (2012). *Alternative rating structures for Kentucky STARS for KIDS NOW, Evaluation Brief #6*. Washington, D.C.: Child Trends.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Education Measurement*, 38, 319-42.
- Kane, M. T. (2006). Validation. In R. Brennan (Ed.) *Educational Measurement, 4th edition* (pp. 17-64). Westport, CT: Praeger.
- Karoly, L. A. and Zellman, G.L. (2012). *How Would Programs Rate Under California’s Proposed Quality Rating and Improvement System? Evidence from Statewide and County Data on Early Care and Education Program Quality*. Santa Monica, CA: RAND Corporation.
- Keyes, C. (2002). A way of thinking about parent/teacher partnerships for teachers. *International Journal of Early Years Education*, 10(3), 177 – 191.
- Kontos, S. (1987). The attitudinal context of family day care relationships. In D. Peters & S. Kontos (Eds.), *Continuity and discontinuity of experience in child care* (pp. 91 – 113). Norwood, NJ: Ablex Publishing.
- Lugo-Gil, J., Sattar, S., Boss, C., Boller, K., Tout, K., & Kirby, G. (2011). *The Quality Rating and Improvement System (QRIS) Evaluation Toolkit. OPRE Report #2011-31*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.
- Manlove, E., Benson, M., Strickland, M., & Fiene, R. (2011). *A comparison of regulated child care in rural and urban Pennsylvania*. Harrisburg, Pennsylvania: The Center for Rural Pennsylvania.
- McGrath, W. (2007). Ambivalent partners: Power, trust, and partnerships in relationships between mothers and teachers in a full-time child care center. *Teachers College Record*, 109(6), 1401 – 1422.
- McWayne, C. & Melzi, G., (2011). Family engagement during preschool, paper presented to the Head Start Advisory Committee on Research and Evaluation, Washington DC.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychology*, 30, 955-966.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Minnesota Department of Education and Minnesota Department of Human Services. (January, 2007). *Child care information and rating system – parent focus group results*. DHS-4965-ENG 1-07. St. Paul, MN.

- Office of Child Development and Early Learning (OCDEL) (2010). *Keystone STARS Program Report*. Harrisburg, Pennsylvania: Department of Public Welfare.
- Pianta, R.C., La Paro, K.M., & Hamre, B.K. (2008). *Classroom Assessment and Scoring System (CLASS)*. Baltimore, MD: Paul H. Brookes Publishing Co, Inc.
- Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131-151.
- Shimoni, R. (1992) Parent involvement in early childhood education and day care, *Sociological Studies of Child Development*, 5, 73–95.
- Tenopyr, M.L. (1977). Content-construct confusion. *Personnel Psychology*, 30, 47-54.
- Thornburg, K., Mayfield, W.A., Hawks, J.S., & Fuger, K.L. (2009). *The Missouri Quality Rating System School Readiness Study*. University of Missouri--Columbia. Center for Family Policy and Research.
- Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the Next Decade of Quality Rating and Improvement Systems*, OPRE Issue Brief. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G. & Boller, K. (2010). *Compendium of Quality Rating Systems and Evaluations*. OPRE Report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Tout, K., Starr, R., Isner, T., Cleveland, J., Albertson-Junkans, L., Soli, M., & Quinn, K. (2011). *Evaluation of Parent Aware: Minnesota's Quality Rating and Improvement System Pilot, Final Evaluation Report*. Produced for the Minnesota Early Learning Foundation. Minneapolis, MN: Child Trends.
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L. & Burchinal, M. (2010). *Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature*, OPRE 2011-5. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zellman, G.L., Brandon, R.N., Boller, K., & Kreader, J.L. (2011). *Effective evaluation of quality rating and improvement systems for early care and education and school-age care, Research-to-Policy, Research-to-Practice Brief* OPRE 2011-11a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zellman, G. L., & Karoly, L.A. (2012). *Incorporating Child Assessments into State Early Childhood Quality Improvement Initiatives*. Santa Monica, CA: RAND Corporation.
- Zellman, G.L., & Karoly, L.A. (2012). *Moving to Outcomes: Approaches to Incorporating Child Assessments into State Early Childhood Quality Rating and Improvement Systems*. Santa Monica, CA: RAND Corporation, OP-364-PF.
- Zellman, G.L. & Perlman, M. (2006). Parent involvement in child care settings: Conceptual and measurement issues. *Early Child Development and Care*, 176(5), 521-538.
- Zellman, G. L., & Perlman, M. (2008). *Child-Care Quality Rating and Improvement Systems in Five Pioneer States: Implementation Issues and Lessons Learned*. Santa Monica, CA: RAND Corporation.
- Zellman, G. L., Perlman, M., Le, V., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar Early Learning quality rating and improvement system as a tool for improving child-care quality*. (MG-650-QEL). Santa Monica, CA: RAND Corporation.

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.

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 Rules¹. 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://dec.al.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

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

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

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

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 3

Key Indicators with Core Rules and Licensing Study	r =	p <	n =
Key Indicators and Core Rules (CCC)	.49	.0001	104
Key Indicators and Licensing Study (CCC)	.78	.0001	104
Key Indicators and Core Rules (FCC)	.57	.0001	147
Key Indicators and Licensing Study (FCC)	.87	.0001	147

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 4

Core Rules with Licensing Studies and ACDW	r =	p <	n =
Core Rules and Licensing Studies (CCC)	.69	.0001	104
Core Rules and ACDW (CCC)	.76	.0001	104
Core Rules and Licensing Studies (FCC)	.74	.0001	147
Core Rules and ACDW (FCC)	.70	.0001	147

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 (See (<http://dec.al.ga.gov/BftS/ResearchStudyOfQuality.aspx>)). 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
Program Compliance and Quality Comparisons

	r =	p <	n=
ECERS-R/PK + ITERS-R and Licensing Studies	.48	.001	45
ECERS-R/PK + ITERS-R and Core Rules	.60	.0001	45
ECERS-R/PS + ITERS-R and Licensing Studies	.21	ns	45
ECERS-R/PS + ITERS-R and Core Rules	.27	ns	45
FCCERS-R and Licensing Studies	.19	.04	146
FCCERS-R and Core Rules	.17	.03	146

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
Environment Rating Scale Scores

	FPG	This Study
ECERS-R Pre-K Total Scale Scores	4.16	4.15
ECERS-R Preschool Total Scale Scores	3.39	3.42
ITERS-R Total Scale Scores	2.74	2.72
FCCERS-R Total Scale Scores	2.50	2.49

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

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

**DMLMA© Expected r Value Thresholds in Order to be Validated (Also see Figure 1 for additional details):*

High correlations (.70+) = Licensing Studies x Key Indicators.

Moderate correlations (.50+) = Licensing Studies x Core Rules; Core Rules x ACDW; Core Rules x Key Indicators; Key Indicators x ACDW.

Lower correlations (.30+) = Program Quality Tools x Licensing Studies; Program Quality x Core Rules; Program Quality x Key Indicators.

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

****ERS = ECERS-R + ITERS-R**

PK = Pre-K program

PS= Preschool program

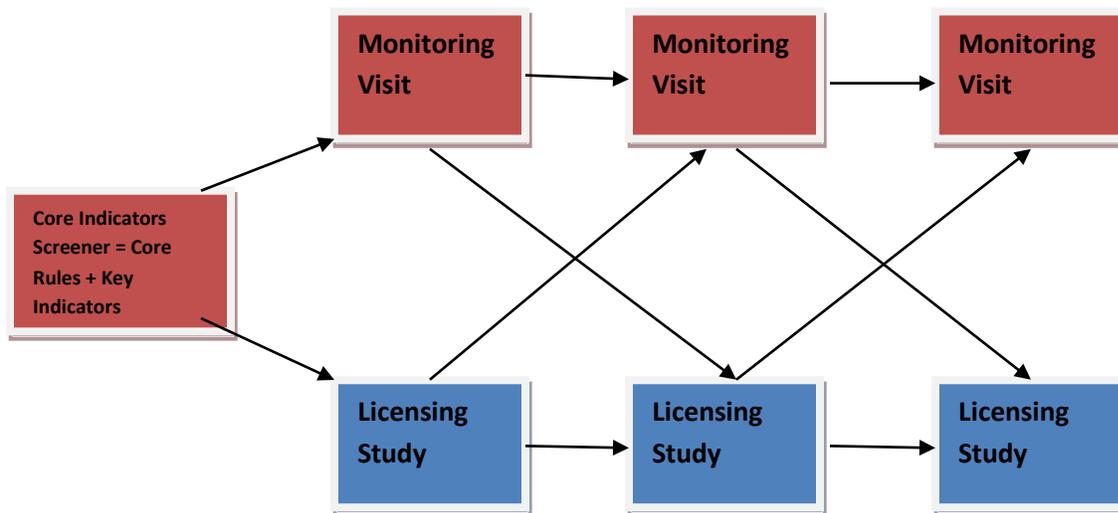
A confounding of data occurred with the first two validation approaches because the Core Rules were influenced a great deal by the National Child Care Key Indicators (NCKKI) (Fiene, 2002) where 10 of the 13 Core Rules overlapped significantly with the NCKKI. This helped to increase the correlation between 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 NCKKI, it should enhance the differential monitoring approach utilized by DECAL.

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 recommendations³ 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:



Compliance Decisions:

Core Indicators = 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).

³ 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

- Fiene (2012). *Differential monitoring logic model and algorithm (DMLMA©)*. Middletown, PA: Research Institute for Key Indicators.
- Fiene (2007). Child Development Program Evaluation & Caregiver Observation Scale, in T Halle (Ed.), *Early Care and Education Quality Measures Compendium*, Washington, D.C.: Child Trends.
- Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, Winter 2002-2003, pps 12-13.
- Fiene (2002). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.
- Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services*, 5(2), 38-39.
- Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice*, 1(2), 10-11.
- Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.
- Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, 14(3), 198-214.
- Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.
- Harms, T., Cryer, D.R., & Clifford, R.M. (2007). Family child care environment rating scale: Revised edition. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2006). *Infant/toddler environment rating scale: Revised edition*. New York: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early childhood environment rating scale: Revised edition*. New York: Teachers College Press.
- Maxwell, K. L., Early, D. M., Bryant, D., Kraus, S., Hume, K., & Crawford, G. (2009a). Georgia study of early care and education: Child care center findings. Chapel Hill, NC: The University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Maxwell, K. L., Early, D. M., Bryant, D., Kraus, S., Hume, K., & Crawford, G. (2009b). *Georgia study of early care and education: Findings from Georgia's Pre-K Program*. Chapel Hill, NC: The University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Maxwell, K. L., Early, D. M., Bryant, D., Kraus, S., & Hume, K., (2010). Georgia study of early care and education: Family child care findings. Chapel Hill: The University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Morgan, Stevenson, Fiene, & Stephens (1986). Gaps and excesses in the regulation of child day care, *Reviews of Infectious Diseases--Infectious Diseases in Child Day Care: Management and Prevention*, 8(4), 634-643.
- Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

TABLE 8 - FOUR APPROACHES TO VALIDATING A QRIS (Zellman & Fiene, 2012)

Approach	Activities and Purpose	Typical Questions Approach Addresses	Issues and Limitations
<i>1. Examine the validity of key underlying concepts</i>	Assess whether basic QRIS quality components and standards are the “right” ones by examining levels of empirical and expert support.	Do the quality components capture the key elements of quality? Is there sufficient empirical and expert support for including each standard?	Different QRISs may use different decision rules about what standards to include in the system.
<i>2. Examine the measurement strategy and the psychometric properties of the measures used to assess quality</i>	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.	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?	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.

TABLE 8 (CONTINUED)

Approach	Activities and Purpose	Typical Questions Approach Addresses	Issues and Limitations
<i>3. Assess the outputs of the rating process</i>	<p>Examine variation and patterns of program-level ratings within and across program types to ensure that the ratings are functioning as intended.</p> <p>Examine relationship of program-level ratings to other quality indicators to determine if ratings are assessing quality in expected ways.</p> <p>Examine alternate cut points and rules to determine how well the ratings distinguish different levels of quality.</p>	<p>Do programs with different program-level ratings differ in meaningful ways on alternative quality measures?</p> <p>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?</p>	<p>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.</p>
<i>4. Examine how ratings are associated with children's outcomes.</i>	<p>Examine the relationship between program-level ratings and selected child outcomes to determine whether higher program ratings are associated with better child outcomes.</p>	<p>Do children who attend higher-rated programs have greater gains in skills than children who attend lower-quality programs?</p>	<p>Appropriate demographic and program level control variables must be included in analyses to account for selection factors.</p> <p>Studies could be done on child and program samples to save resources.</p> <p>Findings do not permit attribution of causality about QRIS participation but inferences can be made about how quality influences children's outcomes.</p>

FIGURE 1- DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (Fiene, 2012)
***DMLMA*© Applied to the Georgia Child Care Licensing System**

$$CI + PQ \Rightarrow RA + KI \Rightarrow DM$$

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 (.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.

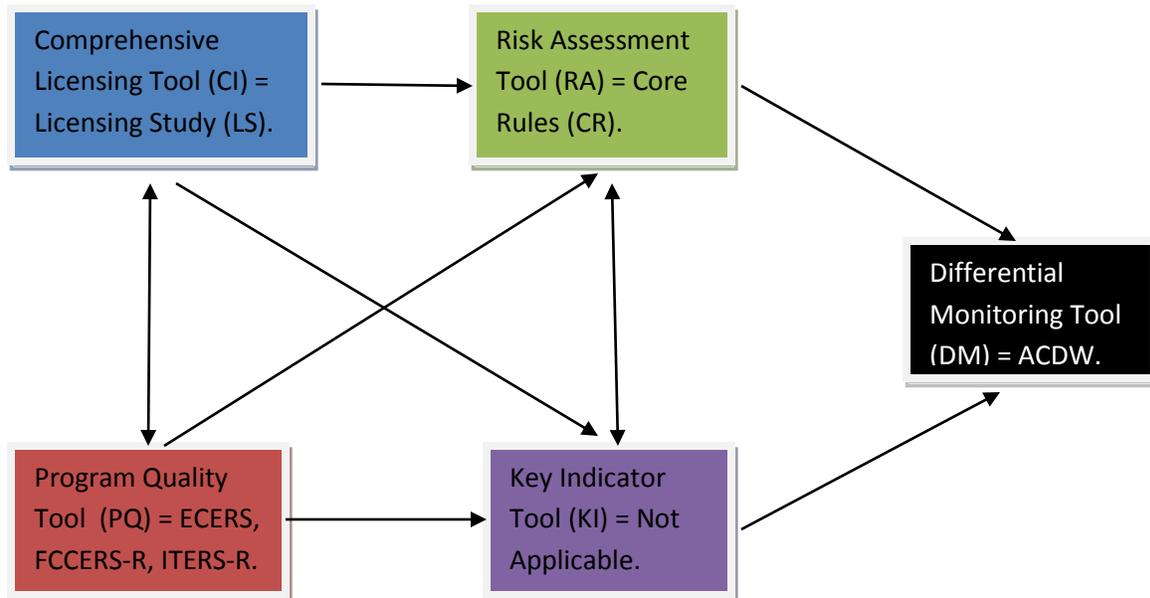


Table 9 - Listing of Key Indicators for Georgia Child Care Centers with Phi Coefficients

591-1-1-25 (3) requires that the center and surrounding premises be clean, free of debris and in good repair. (Phi = .49)

591-1-1-25 (13) requires that hazardous equipment, materials and supplies be inaccessible to children. (Phi = .46)

591-1-1-26 (6) requires that outdoor equipment be free of hazards such as lead-based paint, sharp corners, rust and splinters. (Phi = .44)

591-1-1-26 (8) requires the playground to be kept clean, free of litter and hazards. (Phi = .59)

591-1-1-26 (7) requires that a resilient surface be provided and maintained beneath the fall zone of climbing and swinging equipment. (Phi = .57)

591-1-1-36 (6)(a-c) 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. (Phi = .49)

591-1-1-14 (1) requires that at least 50% of the caregiver staff have current first aid and CPR training. (Phi = .49)

591-1-1-08 (a)-(f) 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.... (Phi = .44)

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

290.2-3-11(2)(C) requires that fire drills be practiced monthly and shall be documented and kept on file for one year. (Phi = .51)

290-2-3-11 (2)(f) requires that poisons, medicines, cleaning agents and other hazardous materials be in locked areas or inaccessible to children. (Phi = .61)

290-2-3-11 (1)(f) requires the family day care home and any vehicle used to have a first aid kit..... (Phi = .57)

290-2-3-07 (4) 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. (Phi = .58)

290-2-3-08 (1)(a) 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. (Phi = .63)

290-2-3-08 (1)(b) 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. (Phi = .57)

290-2-3-08 (1)(b) 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. (Phi = .55)

290-2-3-08 (1)(f) 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. (Phi = .51)

290-2-3-08 (1)(c) 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. (Phi = .72)

Table 11 - Key Indicator Formula Matrix for Generating Key Indicators*

	<i>Providers In Compliance on Rule</i>	<i>Programs Out Of Compliance on Rule</i>	<i>Row Total</i>
<i>High Group**</i>	A	B	Y
<i>Low Group***</i>	C	D	Z
<i>Column Total</i>	W	X	Grand Total

(* This computation occurred for each licensing rule)

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

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

-
- A = High Group + Programs in Compliance on Specific Rule.*
 - B = High Group + Programs out of Compliance on Specific Rule.*
 - C = Low Group + Programs in Compliance on Specific Rule.*
 - D = Low Group + Programs out of Compliance on Specific Rule.*
 - W = Total Number of Programs in Compliance on Specific Rule.*
 - X = Total Number of Programs out of Compliance on Specific Rule.*
 - Y = Total Number of Programs in High Group.*
 - Z = 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

Phi Coefficient Range	Characteristic of Indicator	Decision
(+1.00) – (+.26)	Good Predictor	Include
(+.25) – (-.25)	Unpredictable	Do not Include
(-.26) – (-1.00)	Terrible Predictor	Do not Include

Table 13 - Comparison of the Pre-K and Preschool Programs

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

*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.

Georgia Licensing Validation Technical Elements Appendix

Because of the nature of this report being a state's first attempt at fully validating its 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@decals.ga.gov

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



Contents lists available at [ScienceDirect](#)

Early Childhood Research Quarterly



Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs

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

^a University of Southern Maine, Muskie School of Public Service, PO Box 9300, Portland, ME 04104-9300, United States

^b Purdue University, United States

^c RAND Corporation, United States

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Child care quality
QRIS
Study design
Methodology

ABSTRACT

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

© 2014 Elsevier Inc. All rights reserved.

Introduction

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

and, therefore, improvements to children's cognitive, language, social, emotional, and physical development.

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

Recent research has documented the importance of both specificity and thresholds when testing hypotheses about child care quality impacts on children's developmental outcomes (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Howes, Whitebook, & Phillips, 1992; NICHD ECCRN, 2000, 2002). However, common global measures of classroom quality such as the Early Childhood Environment

* 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.

Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 2005) are not always significantly associated with specific child outcomes (Burchinal, Kainz, & Cai, 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 be 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.

Approach	Activities and purpose	Typical questions	Issue and limitations
1. <i>Examine the validity of key underlying concepts.</i>	Assess whether basic QRIS quality components and standards are the “right” ones to include by examining levels of empirical and expert support.	*Do the quality components capture the key elements of quality? *Is there sufficient empirical and expert support for including each standard?	*Process subject to interpretation and to political pressure *Limited empirical evidence available; few established links to outcomes of interest.
2. <i>Examine the measurement strategy and psychometric properties of measures used to assess quality.</i>	*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 relationships among quality measures to assess whether they function as expected.	*What is the reliability and accuracy of indicators collected using different methods? *Do quality measures perform as expected? (e.g., do expected subscales emerge?) *Do measures of similar standards relate more closely to each other than to other measures? *Do measures relate to each other in theoretically consistent ways?	*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.
3. <i>Assess the outputs of the rating process</i>	*Examine variation and patterns of program-level ratings within and across program types, to assess if QRIS distinguishes levels of quality. *Examine relationship of program-level ratings to other validated quality indicators to determine if ratings are assessing quality in expected ways. *Examine alternate cut points and combining rules to determine how well the ratings distinguish different levels of quality.	*Do rating distributions vary by program type, e.g., center-based programs vs. home-based programs? *Do programs with different program-level ratings differ in meaningful ways on alternative quality measures? *Do levels cut scores and combining rules produce expected rating distributions and meaningful distinctions among programs?	*Measurement error is an important issue that should be examined. *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. *Comparing QRIS measures to other measures is frequently constrained by the absence of validated alternative measures of the same constructs.
4. <i>Examine how ratings are associated with children's outcomes.</i>	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 are attending higher-rated programs have greater levels or 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 may be done on selected child and program samples to save resources. *Given staff turnover and movement, children may not spend substantial time with quality-rated caregivers (dosage.) *Measurement limitations noted for Approach 3 apply here as well. *Findings from non-experimental studies do not permit attribution of causality about QRIS participation, but preliminary inferences can be made about how quality influences children's outcomes.

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

Both states partnered with university-based researchers to conduct validation research, after piloting aspects of their QRIS design. However, there are also key differences between these two states. For example, the Indiana QRIS standards were developed based on a local community-based model that was then modified by a state stakeholder committee for statewide expansion. The Maine quality standards were developed to align with program-type-specific national accreditation standards. The Maine

standards were also vetted through review and comment by many stakeholders and technical assistance was provided by University researchers based on reviews of the scientific literature. Maine QRIS ratings are generated by provider self-report, then verified by state agency staff, while Indiana employs independent raters who directly assess the standards by visiting child care settings. Provider voluntary participation rates are higher among state-licensed providers in Indiana. However, Indiana also has significant numbers of license-exempt child care providers, whereas license exemption is not a prominent feature of the Maine child care system. The key features of each state QRIS are summarized in Table 2. These two states provide useful examples, because while the state child care contexts are different, they each used strategies contained in the four validation approaches discussed above and outlined in Table 1. The successes and limitations of these states' approaches will inform future validation research on QRIS.

Table 2
Key features of Indiana and Maine QRISs.

QRIS feature	Indiana	Maine
System name	Paths to QUALITY (2008)	Quality for ME (2008)
Eligible child care types (participation rate)	Licensed centers (89%) Licensed homes (62%) Unlicensed registered ministry centers (12%)	Licensed centers (68%) Licensed homes (44%)
Participation rules	Voluntary for all providers (Not required for CCDF participation)	Required for programs participating in federal CCDF/Voluntary for all others
QRIS structure	Building Block; 4 quality levels	Building Block; 4 quality levels
QRIS standards (examples)	Level 1 – Licensed, or completes voluntary certification program Level 2 – Learning environment and materials requirements; daily literacy activities; 25% of staff have CDA or equivalent; 15 hrs. in-service training/yr.; etc. 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. Level 4 – National accreditation; Provide mentoring to other QRIS providers (see www.in.gov/fssa/carefinder/2554.htm#).	Step 1 – Meets all regulatory standards, in operation for more than one year, and all staff registered in Maine Roads to Quality Registry (MRTQ). Step 2 – Learning Environment/Developmentally Appropriate Practice requirements; program improvement plan in place; 50% of staff at level 5 on MRTQ career lattice; etc. 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. Step 4 – National accreditation; written parent involvement plan; etc. (See: https://www.maine.gov/dhhs/ocfs/ec/occhs/qualityforme.htm).
QRIS standards development process	Aligned with national accreditation standards, expansion of community pilot program, modified by state stakeholder committee.	Aligned with national accreditation standards; based on available empirical evidence; and from results of stakeholder reviews and comment.
QRIS rating procedure	Independent ratings contractor, annual site visits, using Paths to QUALITY standards checklist	(1) Provider self-assessment, online system; (2) enrollment system uses linked files from state licensing and registry; and (3) verified by state agency staff.

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, & Zaslow, 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 rated as follows: (1) some or limited evidence (one or two well-designed studies); (2) moderate evidence (3–5 well-designed studies); or (3) substantial evidence (more than five well-designed studies). The results of this analysis were reported to the state QRIS planning committee, including a conclusion that most of the proposed quality indicators had “substantial evidence” for their validity.

Maine: examining the validity of underlying concepts

Researchers at the University of Southern Maine worked with state agency leaders and other key stakeholders through a process that involved the use of Concept Mapping ([The Concept System®](#), 2012). This process allows for the development of a conceptual framework that can guide planning, and in this case led to the selection of the underlying quality concepts and standards for Maine's QRIS. Similar to what was done in Indiana, University of

Southern Maine research staff identified key quality concepts from the literature and national accreditation standards. In addition, concepts emerged from results of eight focus groups with parents and ECEC professionals across the state, including participants from various types of settings, e.g., family child care homes, after school programs, centers, etc. Statements of program quality were developed; these statements were the focus of a mapping process which involved more than twenty-four experts reviewing and rating the statements. The Concept Maps that result from this process allowed participants to visually identify which concepts of program quality were most favored by specific key stakeholder groups. In addition, the mapping software illustrated how closely related the concepts were to each other, based on reviews from the select experts. From this process, a set of components and standards was developed. The final step in selecting program quality standards involved a formal review and comment process that the state agency implemented in various locations across the state ([Maine DHHS, 2008](#)).

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

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

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

At the time both the Maine and Indiana QRISs were being designed, in the mid-2000s, the predominant global quality measures in use in both states were the Environmental Rating Scales (ERS) ([Harms et al., 2005](#); [Harms, Cryer, & Clifford, 2006](#); [Harms, Cryer, & Clifford, 2007](#)). ERS use was predominant in the accreditation quality improvement efforts in both child care centers and child care homes. So there was some familiarity with the measures on the part of providers. This was an important political consideration in terms of developing and promoting the design of the QRIS. In addition, in reviewing emerging QRIS work from other states, it appeared that the ERSs were the predominant global classroom quality measure in use at that time. While ERS was influential in the design of the QRIS quality standards in both Maine and Indiana, it is important to note that the ERS are not used to determine the step or level quality ratings. Many other quality indicators are included in the QRIS standards of both states, including staff qualifications, annual staff training hours, and other indicators that help providers make progress toward the ultimate quality goal of national accreditation.

Choosing the points at which individual measures (in block design QRISs) and summary ratings are assigned to rating levels is another exercise that has received limited attention. Cut scores can be assessed in a number of ways. One relatively simple one is to use existing data to conduct a “virtual pilot” (Zellman & Karoly, 2012a) in which existing data are used and cut scores are altered and the effects are examined in terms of distributions of summary ratings across programs. A downward limit on cut scores is the need for some variation within each quality component; without it, a component provides no useful information in overall ratings. Designers may compare program distributions using different cut scores, although it is not always clear what an appropriate rating levels distribution should be. However, it is reasonable to assume that an appropriate distribution in the early phase of a QRIS would be one in which there are programs placed at all levels, with decreasing numbers of programs at each succeeding higher level.

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

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

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

Approach 3: assess the outputs of the rating process

A third validation approach focuses on assessing the outputs of the rating system: the scores and levels assigned to providers who undergo a rating, and the distributions of those scores within and across different types of providers. Studies conducted under this approach examine the degree to which the quality levels in the QRIS are meaningfully distinct from each other. The results of these

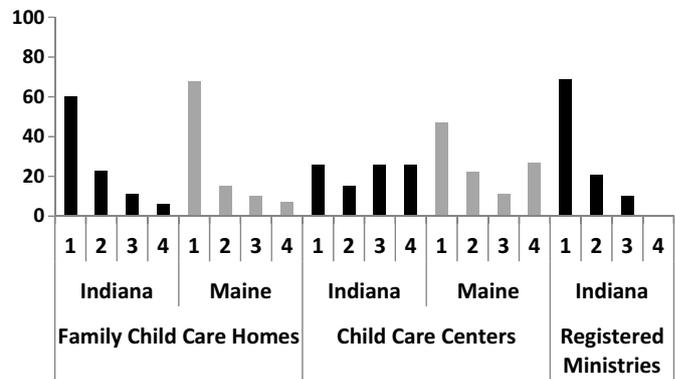


Fig. 1. Percent enrolled providers rated at four quality levels, by state and type of child care.

studies may provide data that suggest that measures, cut scores, or rules for combining measures need to be changed in order to distinguish the rated quality levels effectively. Because these studies can result in proposals for significant changes to the standards for QRIS levels, it is helpful for these studies to occur prior to studies that examine associations between quality levels and children’s development.

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

Examining initial QRIS rating distributions and cut points

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

In Indiana, licensed child care centers were evenly distributed across the four QRIS levels approximately two years after the program inception. However licensed family child care homes were most frequently found at Level 1, with steeply declining numbers at the other three levels. This higher proportion of Indiana licensed centers rated at Level 3 or Level 4 may have been due to a greater historical emphasis in child care centers than in homes on regulation and attaining national accreditation, greater organizational capacity to complete the requirements of advancement in QRIS, or possibly that QRIS standards more closely reflect center quality than family child care home quality. Unlicensed registered child care 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^a as a function of program type and rated quality level.

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

^a Possible range = 1–7.

Table 4
Maine QRIS: mean global quality ERS scores^a as a function of program type and rated quality level.

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

^a 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 ministry centers had advanced at least one QRIS quality level in a 6-month period between assessments, during which mentoring was provided by local training providers. This advancement pattern, if maintained over time, suggests that even though attaining the highest levels may be challenging, quality improvement is feasible.

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

An assumption of the designers of the QRIS in Maine was that programs engaged with QRIS will improve their tier levels consistently over time (Lahti et al., 2011). Approximately 80% of all programs (n=1118) in the QRIS observed during the study period 2008 through 2011 did not experience a move up from one Step Level to the next. Results indicated that 95 of the 103 events or changes in Step Level from level one to two occurred during the first 23 months of enrollment. Moving from a Step One to Two, center-based care programs had a hazard probability of just .02 while family child care homes stayed virtually flat during this early period of enrollment in the QRIS. For movement from Step Two to Three, neither program type (p=290) nor regional location (p=195) appear to be significant in explaining Step level movement. For movement from Step Three to Four, the highest tiers in Maine’s QRIS, the analysis indicated that only type of program is a significant covariate explaining advancement. Family child care homes

appeared to have a significantly lower probability of advancing a Step Level at this highest quality tier, compared with center-based and Head Start programs. These types of analyses of program movement in the system are relevant to the validation of a QRIS as they illustrate whether or not the way the system as designed is meeting its goals of supporting program advancement, leading to statewide improvement of program quality.

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

Indiana: assessing the output of the rating process

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

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

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

Table 5
Indiana QRIS: mean total positive interaction (CIS) scores^a as a function of program type and rated quality level.

	Level one (n = 84)	Level two (n = 87)	Level three (n = 74)	Level four (n = 65)
All providers (N = 310)	3.0 (.44)	3.1 (.46)	3.2 (.37)	3.2 (.52)
Family child care homes (n = 164)	2.9 (.37)	3.1 (.43)	3.2 (.36)	3.3 (.48)
Licensed child care centers (n = 132)	3.2 (.44)	3.1 (.50)	3.2 (.41)	3.2 (.54)
Unlicensed registered child care ministries (n = 14)	2.8 (.58)	3.2 (.48)	3.5 (.11)	NA

^a Possible range = 1–4.

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 output of the rating process

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

The results show an overall significant difference between Step Level and ERS mean score at the classroom/setting level ($F = 5.02$; $df = 3, 307$; $p = .002$). Results of post hoc Bonferroni tests showed a significant difference between Step One and Step Four programs ($p = .001$) and between Step Two and Step Four programs ($p = .001$). The total variance of the mean ERS score explained by Step Level was only 5%, indicating weak relationships between the variables. Comparisons of the program type mean ERS quality scores indicated a difference only between the family child care 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 Schultze (2000), and was found not to be correlated to Step Level rating (Pearson's $r = .010$, $p = .68$). In terms of services parents should have received according to program standards, a majority of parents reported not receiving: information about other government services for their child; opportunities for parent engagement with the program; daily communication from the program about their child; and being provided an up to date written parent hand-book from their provider. There did not appear to be any difference in step level in terms of parents not consistently receiving these types of services as required by the program quality standards according to parent reports. The use of these data by QRIS administrators was primarily for monitoring purposes focused on services and or supports parents should have received based on requirements in the QRIS standards. While parents were asked about perceptions of program quality, due to a strong desire to reach out to parents as a key stakeholder in the QRIS, that information was not relied upon for program planning or program improvement.

Approach 4: relate ratings to children's development

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

The logic models that underlie QRISs typically assert that higher quality care will be associated with improved child outcomes. Therefore, one important piece of validation evidence concerns whether children make greater developmental gains in programs with higher program-level QRIS ratings than in programs with lower ratings. While a definitive evaluation of QRIS impact on child outcomes would consist of an experimental study with random assignment of providers and children to QRIS levels, ethical and practical considerations often make experiments impractical, at least on a state-level scale. Instead, current studies evaluating QRIS validity in terms of child outcomes using this approach do not attempt to evaluate causal linkages. Instead, they examine whether the QRIS ratings and the quality components that comprise the ratings are associated in expected ways to measures of children's development. Showing significant associations between QRIS-rated quality would be a first step, a necessary but not sufficient result to demonstrate causal inferences about how QRIS quality influences children's outcomes.

To date, few QRIS validation studies have incorporated children's outcomes. Maine did not include this approach to validation. As Elicker and Thornburg (2011) note, results from such studies are mixed, at least in part because of the challenges of conducting them. A primary challenge is the inability to control for all the factors that may confound the quality-outcome correlations for children whose families have selected programs in a non-random way. Additional challenges include the difficulty of recruiting of programs and children across all quality levels; lack of information about the amount of care children received in each setting (dosage); lack of appropriate outcome measures for children of diverse ages, abilities, cultures and linguistic backgrounds; and, lack of variation in the quality of participating QRIS programs. As noted above in the discussion of Approach 3, measurement error remains a problem.

Indiana: examination of ratings associated with children's outcomes

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

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

Bivariate correlations and multiple regression models were used to explore the associations between children's development and the three measures of child care quality: QRIS ratings (4 levels); ERS global quality scores; and a CIS composite rating of positive adult-child interactions. All regression models included parent education, household income, and type of child care. No significant correlations were found between the four-level QRIS quality ratings and either infant/toddler or preschooler developmental status. Some of the researcher-observed quality measures were mildly but significantly correlated with child development measures. For preschoolers, CIS positive interactions were correlated with social competence ($r = .17^{**}$) 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 of Indiana and Maine). Use of these approaches enables cross system comparisons,

which will allow for the identification of common threats to validity and useful strategies to enhance the validity of a state QRIS.

Validation and early care and education system constraints

A QRIS is not merely a program-level quality-improvement intervention, it is a policy lever for strengthening a state's overall early care and education system that reaches beyond child care (Schaack, Tarrant, Boller, & Tout, 2012). The two state validation efforts highlighted in this paper reflect the challenges and constraints common to other state experiences with validation activities (Lahti, Sabol, Starr, Langill, & Tout, 2013). On-site observations of global program quality and establishing and maintaining inter-rater reliability for QRIS raters is a time consuming and costly endeavor. Keen interest in school readiness may pressure program administrators to collect child outcome level data before a QRIS is well established. Current research on the measures that are in the widest use to predict child outcomes 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 QRIS (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

- Alkon, A. (2013). *National health and safety tool*. San Francisco, CA: California Child Care Health Program.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541–552. [http://dx.doi.org/10.1016/0193-3973\(89\)90026-9](http://dx.doi.org/10.1016/0193-3973(89)90026-9)
- Berry, D. J., Bridges, L. J., & Zaslow, M. J. (2004). *Early childhood measures profile*. Retrieved from: <http://aspe.hhs.gov/hsp/ECMeasures04/report.pdf>
- Briggs-Gowan, M. J., & Carter, A. S. (2002). *Brief infant-toddler social and emotional assessment (BITSEA) manual, version 2.0*. New Haven, CT: Yale University.
- Bryant, D. M. (2010). *Observational measures of quality in center-based early care and education programs (OPRE Research-to-Policy, Research-to-Practice Brief 2011-10c)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Bryant, D. M., Burchinal, M., & Zaslow, M. (2011). Empirical approaches to strengthening the measurement of quality: Issues in the development and use of quality measures in research and applied settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 33–47). Baltimore, MD: Paul H. Brookes.
- Burchinal, M. R., Peisner-Feinberg, E. S., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality. Testing for differential associations related to poverty, gender, or ethnicity. *Journal of Applied Developmental Science*, 4, 149–165. http://dx.doi.org/10.1207/S1532480XADS0403_4
- Burchinal, M. R., Vandergrift, N., Pianta, R., & Mashburn, A. J. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in prekindergarten programs. *Early Childhood Research Quarterly*, 25(2), 166–176. <http://dx.doi.org/10.1016/j.ecresq.2009.10.004>
- Burchinal, M. R., Kainz, K., Cai, Y., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2009). *Early care and education quality and child outcomes (Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief #1)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services, and Child Trends.
- Burchinal, M. R., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–27). Baltimore, MD: Paul H. Brookes.
- Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, 25(2), 140–165. <http://dx.doi.org/10.1016/j.ecresq.2009.11.001>
- Cizek, G. J. (2007, August). Introduction to validity. In *Presentation to the National Assessment Governing Board of NAEP* Washington, DC.
- Clifford, R. (2005). Structure and stability in the early childhood environment rating scale. In H. Schonfeld, S. O'Brien, & T. Walsh (Eds.), *Question of quality: Proceedings of a conference on defining, assessing and supporting quality in early childhood care and education* (pp. 12–21). Dublin, Ireland: Centre for Early Childhood Development and Education, St. Patrick's College.

- Cochran, M. (2007). *Finding our way: The future of American early care and education*. Washington, DC: Zero to Three.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service.
- Elicker, J., Langill, C. C., Ruprecht, K., & Kwon, K. (2007). *Paths to QUALITY™: A child care quality rating system for Indiana. What is its scientific basis?* Lafayette, IN: Purdue University, Center for Families.
- Elicker, J., & Thornburg, K. (2011). *Evaluation of quality rating and improvement systems in early childhood programs and school age care: Measuring children's development (Research to Policy, Research to Practice Brief, OPRE 2011-11c)*. Washington, DC: Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.
- Elicker, J., Langill, C. C., Ruprecht, K., Lewsader, J., & Anderson, T. (2011). *Evaluation of "Paths to QUALITY" Indiana's child care quality rating and improvement system*. West Lafayette, IN: Purdue University Center for Families.
- Elicker, J., Langill, C., Ruprecht, K. M., Lewsader, J., Anderson, T., & Brizzi, M. (2013). Paths to QUALITY: Collaborative evaluation of a new child care quality rating and improvement system. *Early Education and Development*, 24(1), 42–62. <http://dx.doi.org/10.1080/10409289.2013.736127>
- Emlen, A., Koren, P., & Schultze, K. (2000). *A packet of scales for measuring quality from a parent's view*. Portland, OR: Portland State University, Regional Research Institute for Human Services, The Oregon Child Care Research Partnership.
- Fiene, R. (2013). *Differential monitoring logic model and algorithm (DMLMA): A new early childhood program quality indicator model (ECPQM) for early care and education regulatory agencies*. Middletown, PA: Research Institute for Key Indicators.
- Fiene, R., & Nixon, M. (1985). Instrument based program monitoring and key indicators in child care. *Child Care Quarterly*, 14(3), 198–214.
- Fuller, B., Loeb, S., Kagan, S. L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type. *Quality, and Stability, Child Development*, 75(1), 47–65.
- Harms, T., Clifford, R., & Cryer, D. (2005). *Early childhood environmental rating scale—revised*. New York, NY: Columbia University Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. (2006). *Infant toddler environmental rating scale—revised*. New York, NY: Columbia University Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. (2007). *Family day care environmental rating scale—revised*. New York, NY: Columbia University Teachers College Press.
- Howes, C., Whitebook, M., & Phillips, D. (1992). Results on the National Child Care Staffing Study concerning effectiveness and characteristics of teachers across the United States. *Child & Youth Care Forum*, 21(6), 399–414.
- Howes, C., Burchinal, M., Pianta, R. C., Bryant, D. M., Early, D. M., Clifford, R. M., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Child & Youth Care Forum*, 37(1), 27–50. <http://dx.doi.org/10.1016/j.ecresq.2007.05.002>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Education Measurement*, 38, 319–342.
- Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly*, 28, 645–657. <http://dx.doi.org/10.1016/j.ecresq.2013.05.001>
- LaFreniere, P. J., & Dumas, J. E. (1997). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment*, 8(4), 369–377. <http://dx.doi.org/10.1037//1040-3590.8.4.369>
- Lahti, M., Cobo-Lewis, A., Dean, A., Rawlings, S., Sawyer, E., & Zollitsch, B. (2011). *Maine's quality for ME – child care quality rating and improvement system (QRIS): Final evaluation report*. Portland, ME: Cutler Institute for Health and Social Policy, Muskie School of Public Service, University of Southern Maine.
- Lahti, M., Sabol, T., Starr, R., Langill, C., & Tout, K. (2013). *Validation of quality rating and improvement systems (QRIS): Examples from four states (Research-to-Policy, Research-to-Practice Brief OPRE 2013-036)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Lugo-Gil, J., Sattar, S., Boss, C., Boller, K., Tout, K., & Kirby, G. (2011). *The quality rating and improvement system (QRIS) evaluation toolkit (OPRE Report #2011-31)*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.
- Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service, Inc.
- NICHD Early Child Care Research NICHD Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Sciences*, 4, 116–135.
- NICHD Early Child Care Research NICHD Network. (2002). Child care structure → process → outcome: Direct and indirect effects of child care quality on young children's development. *Psychological Science*, 13, 199–206.
- Maine, Department of Health and Human Services. (2008). *Quality for ME: Child care quality rating system program standards*. Retrieved from: <http://www.maine.gov/dhhs/ocfs/ec/occhs/quality.cbc.pdf>
- Mathematica Policy Research. (2011). *Defining and measuring quality: An in-depth study of five child care quality rating and improvement systems (OPRE Report 2011-29)*. Retrieved from: <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2011/assets/management.pdf>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual, Pre-K*. Baltimore, MD: Paul H. Brookes.
- QRIS National Learning Network (2014). Retrieved from: <http://qrisnetwork.org/qris-state-contacts-map>
- Schaack, D., Tarrant, K., Boller, K., & Tout, K. (2012). Quality rating and improvement systems, frameworks for early care and education systems change. In S. L. Kagan, & K. Kauerz (Eds.), *Early childhood systems, transforming learning* (pp. 71–86). New York, NY: Teachers College Press.
- Smith, S., Robbins, T., Stagman, S., & Kreader, J. L. (2012). *Practices for promoting young children's learning in QRIS standards*. Retrieved from: <http://www.nccp.org/publications/pdf/text.1070.pdf>
- The Concept System® (2012). BT computer software. PN Concept Systems, Incorporated, PL Ithaca, NY.
- Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the next decade of quality rating and improvement systems, OPRE Issue Brief*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality ratings systems and evaluations*. Washington, DC: Child Trends and Mathematica Policy Research, for Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.
- U.S. Department of Health and Human Services, Administration for Children and Families. (2011). *Program instruction (Child Care and Development Block Grant Act of 1990)*. Retrieved from: <http://www.acf.hhs.gov/sites/default/files/occ/pi2011.03.pdf>
- Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (2011). *Quality measurement in early childhood settings*. Baltimore, MD: Paul H. Brookes.
- Zellman, G. L., & Fiene, R. (2012). *Validation of quality rating and improvement systems for early care and education and school-age care (Research-to-Policy, Research-to-Practice Brief OPRE 2012-29)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zellman, G. L., Perlman, M., Le, V., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar Early Learning quality rating and improvement system as a tool for improving child-care quality (MG-650-QEL)*. Santa Monica, CA: RAND Corporation.
- Zellman, G. L., Brandon, R. N., Boller, K., & Kreader, J. L. (2011). *Effective evaluation of quality rating and improvement systems for early care and education and school-age care (Research-to-Policy, Research-to-Practice Brief OPRE 2011-11a)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zellman, G. L., & Karoly, L. A. (2012). *Incorporating child assessments into state early childhood quality improvement initiatives*. Santa Monica, CA: RAND Corporation.

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).

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

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 level¹.

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

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

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

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 the data are more normally distributed² 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 compliance³. 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.

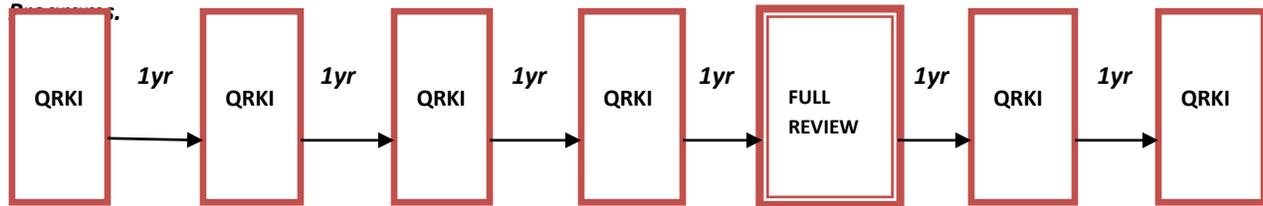
Figure 5 – Thresholds for the Phi Coefficient (Fiene & Nixon, 1983, 1985)(Fiene, 2014)

Phi Coefficient Range	Characteristic of Indicator	Decision
(+1.00) – (+.76)	Good Predictor	Include
(+.75) – (-.25)	Unpredictable	Do not Include
(-.26) – (-1.00)	Terrible Predictor	Do not Include

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.

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

Use of Qualistar Rating Key Indicators (QRKI) for Monitoring with a Full Review every 4th Year for Star 4



This model is taken from the licensing literature and as will be pointed out in the Limitations and Conclusion Sections may not necessarily be appropriate for QRS/QRIS systems depending on a state's QRS/QRIS data distribution. It is provided for illustrative purposes.

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 quality⁴. 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 designation⁵ 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

	Phi	Significance
Family Partnership Standard/Criterion 5	1.00	.001
Family Partnership Standard/Criterion 7	0.86	.001
Family Partnership Standard/Criterion 8	0.83	.001

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 skewed⁶ 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 their 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:

- Barnard, Smith, Fiene, Swanson (2006). Evaluation of Pennsylvania's Keystone STARS Quality Rating and Improvement System, Pittsburgh: Pennsylvania, Office of Child Development.
- Fiene (2013a). A comparison of international child care and US child care using the Child Care Aware – NACCRRRA (National Association of Child Care Resource and Referral Agencies) child care benchmarks, *International Journal of Child Care and Education Policy*, 7(1), 1-15.
- Fiene (2013b)(2014). *Differential monitoring logic model and algorithm*. Middletown: Pennsylvania, Research Institute for Key Indicators.
- Fiene (2013c). Head Start Key Indicators. Middletown: Pennsylvania, Research Institute for Key Indicators.
- Fiene (2013d). Kansas Child Care Key Indicators. Middletown: Pennsylvania, Research Institute for Key Indicators.
- Fiene (2013e). Validation of Georgia's core rule differential monitoring system. Middletown: Pennsylvania, Research Institute for Key Indicators.
- Fiene (2007). Child Development Program Evaluation & Caregiver Observation Scale, in T Halle (Ed.), *Early Care and Education Quality Measures Compendium*, Washington, D.C.: Child Trends.
- Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, winter 2002-2003, pps 12-13.
- Fiene (2002a). *Thirteen indicators of quality child care: Research update*. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services.
- Fiene (2002b). Improving child care quality through an infant caregiver mentoring project, *Child and Youth Care Forum*, 31(2), 75-83.
- Fiene, Iutcovich, Johnson, & Koppel (1998). Child day care quality linked to opportunities for professional development: An applied community psychology example. *Community Psychologist*, 31(1), 10-11.
- Fiene (1996). Using a statistical-indicator methodology for accreditation, in *NAEYC Accreditation: A Decade of Learning and the Years Ahead*, S. Bredekamp & B. Willer, editors, Washington, D.C.: National Association for the Education of Young Children.
- Fiene (1995). Utilizing a statewide training system to improve child day care quality: The other system in a program quality improvement model. *Child Welfare*, Volume LXXIV, #6, November-December, 1189-1201.

- Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services*, 5(2), 38-39.
- Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice*, 1(2), 10-11.
- Fiene, Greenberg, Bergsten, Carl, Fegley, & Gibbons (2002). *The Pennsylvania early childhood quality settings study*, Harrisburg, Pennsylvania: Governor's Task Force on Early Care and Education.
- Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.
- Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly*, 14(3), 198-214.
- Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.
- Kontos & Fiene (1987). Child care quality, compliance with regulations, and children's development: The Pennsylvania Study, in *Quality in Child Care: What Does Research Tell Us?* Phillips, editor, Washington, D.C.: National Association for the Education of Young Children.
- Zellman, G. L. and Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care*, Research-to-Policy, Research-to-Practice Brief OPRE 2012. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services

Appendix – Figure 8

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*)

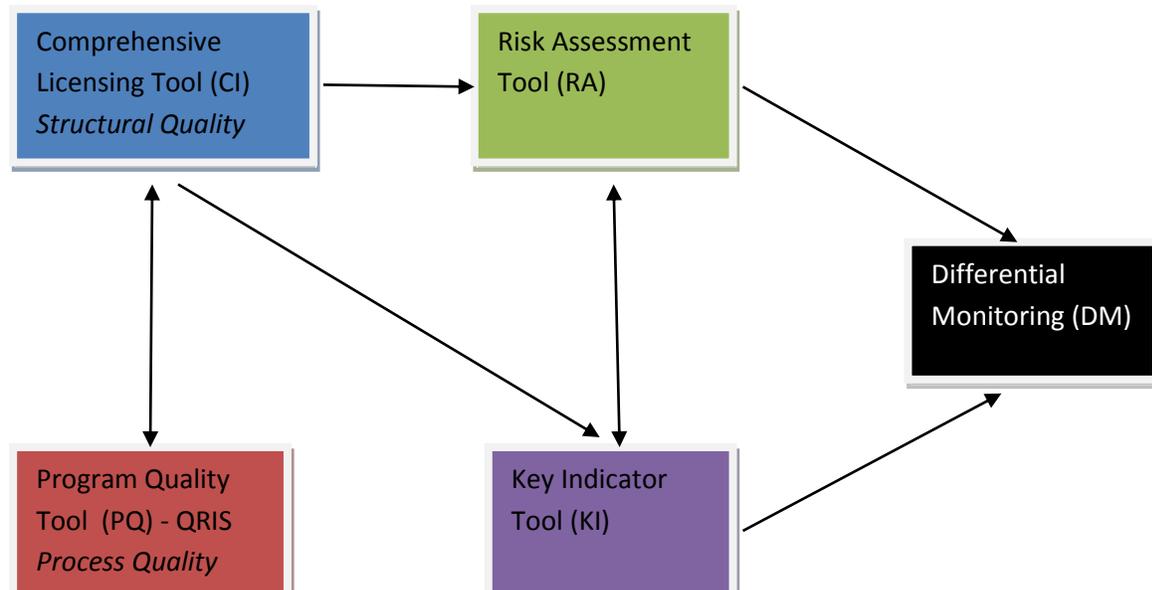
RA = Risk Assessment, (High Risk Rules) (*Stepping Stones*)

KI = Key Indicators (Predictor Rules) (*13 Key Indicators of Quality Child Care*)

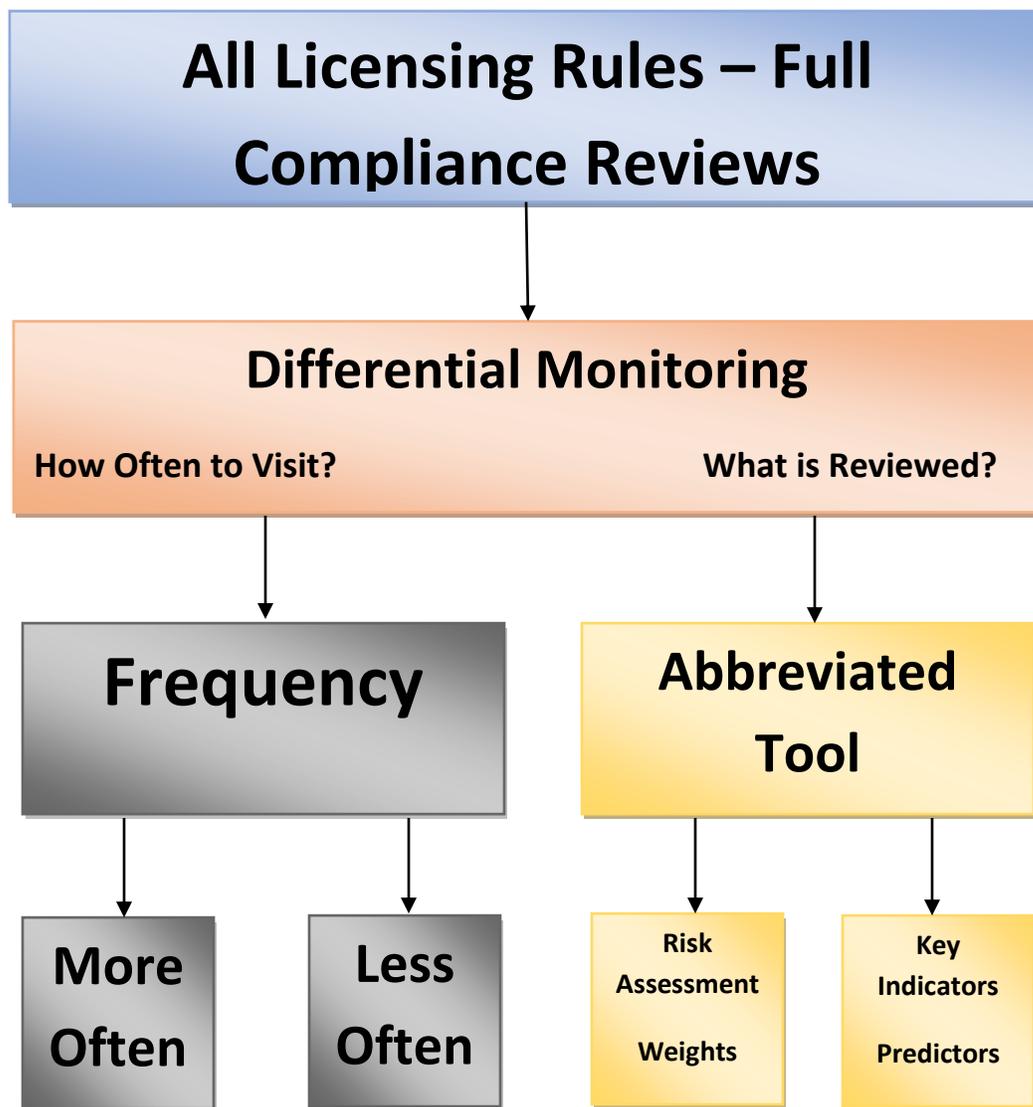
DM = Differential Monitoring (How often to visit and what to review)

PD = Professional Development/Technical Assistance/Training (Not pictured but part of Model)

CO = Child Outcomes (Not pictured but part of Model)

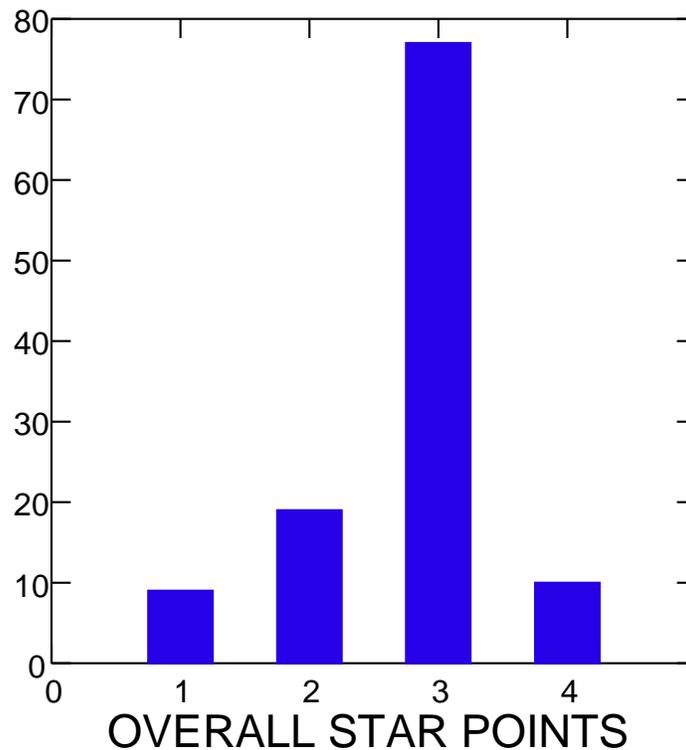


Appendix – Figure 9 - Licensing Rules, Compliance Reviews, Differential Monitoring, Abbreviated Tools, Risk Assessment, and Key Indicators



APPENDIX

Figure 10



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

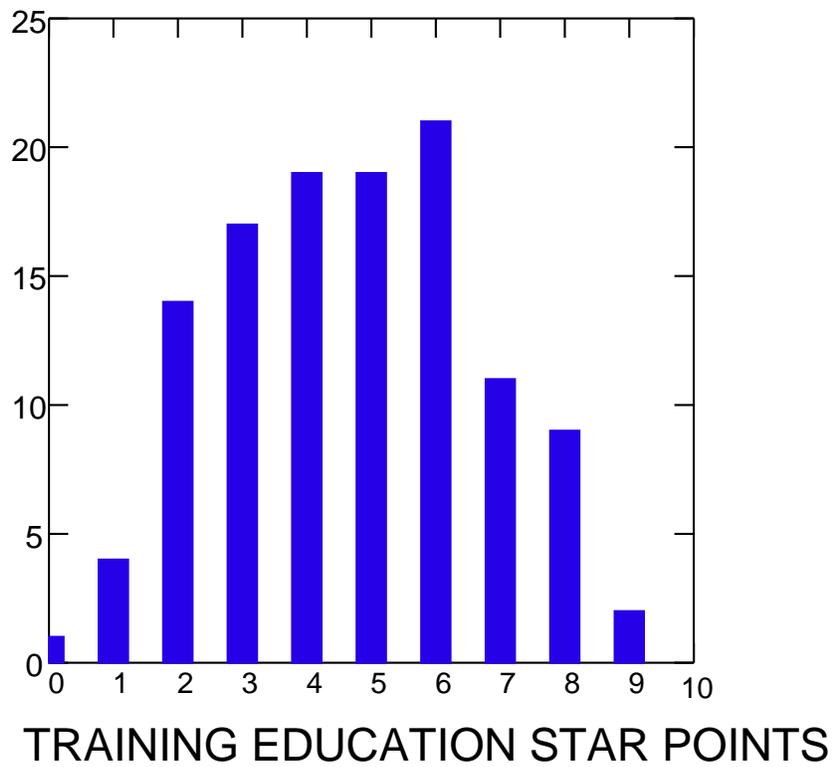


Figure 12

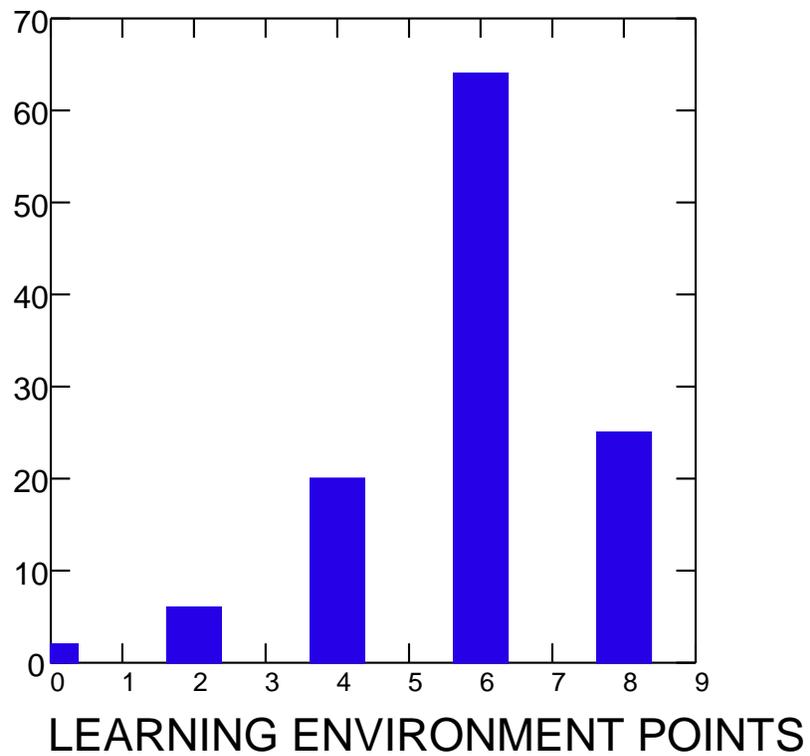


Figure 13

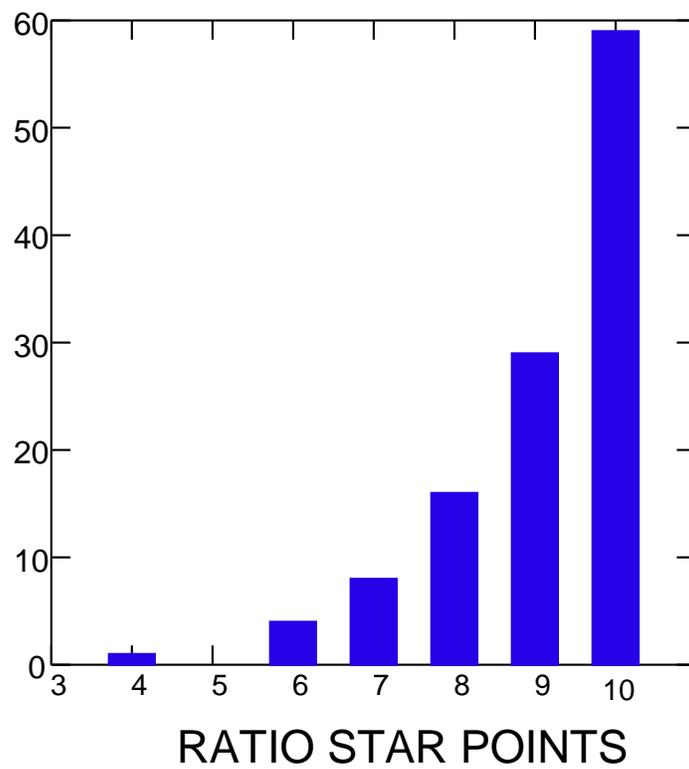


Figure 14

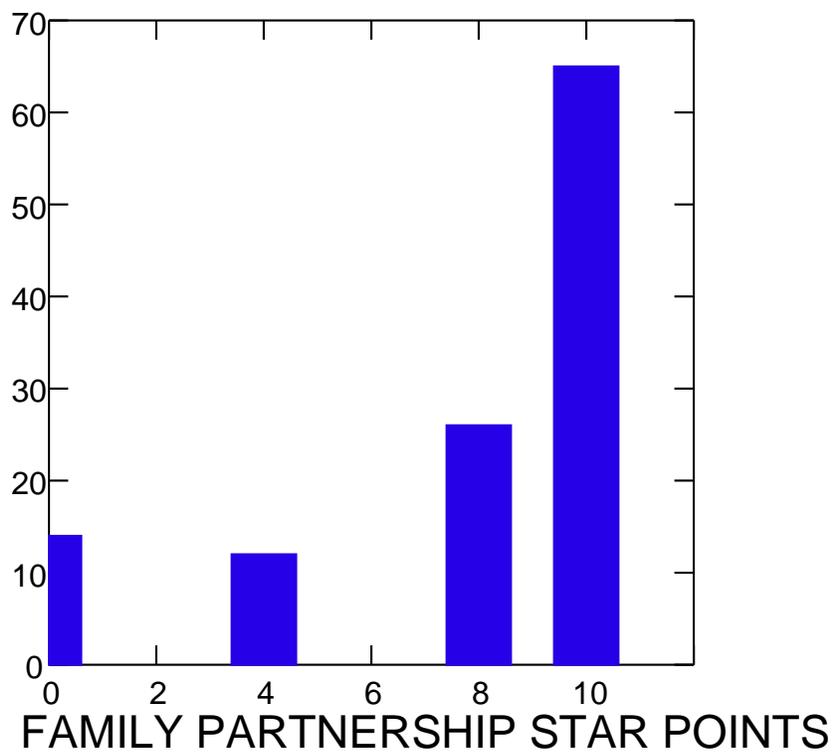


Figure 15

Selected Relationships amongst the Standards/Criteria and Star Level

Standards/Criteria	Correlation (r)
Family Partnerships x Star Level	.80****
Learning Environment x Star Level	.68***
Training/Education x Star Level	.54**
Adult-Child Ratio/Group Size x Star Level	.46*
Program Accreditation x Star Level	.11

* p < .05

** p < .01

*** p < .001

**** p < .0001

Figure 15a

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

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

<u>Adult-Child Ratio/Group Size</u>	<u>Phi</u>	<u>Significance</u>
Adult-Child Ratios	.58	.0001
Group Size	.33	.02

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

Basic Descriptive Statistics for Criteria

<u>Criteria</u>	<u>Mean</u>	<u>Median</u>	<u>Skewness</u>
Family Partnerships	7.7	10	-1.425
Adult-to-Child Ratios & Group Size	9.1	10	-1.506
Learning Environment	5.8	6	-0.946
Training and Education	4.7	5	0.028
Program Accreditation	0.0	0	7.548
Total Star Level	2.7	3	-1.213

Improving Child Care Quality Through an Infant Caregiver Mentoring Project

Richard Fiene

The Pennsylvania State University

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

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

Introduction

This paper describes a child care mentoring project designed to improve the quality of infant and toddler child care programs in south central Pennsylvania. The goal of the mentoring project was to improve the quality of the child care environment and specifically the quality of caregiver-child interactions. As most caregivers in Pennsylvania only receive workshop training, the goal of this project was to compare the mentoring approach to the more typical workshop training. Mentoring is being explored because of its targeted intensive one-on-one nature in delivering training to caregivers based upon needs assessments. The project was conducted during the later half of 2000 and the beginning of 2001. The results presented in this paper are part of the pre- and post-test data collection phase (summer 2000 and winter 2000–2001) of this mentoring project. The actual mentoring intervention occurred from September through December 2000.

Correspondence should be directed to Richard Fiene, Ph.D., Director, Capital Area Early Childhood Training Institute, and Senior Research Associate, Prevention Research Center, The Pennsylvania State University, 2001 North Front Street, Suite 314, Harrisburg, PA 17102; e-mail: rjf8@psu.edu, Web page: <http://caecti.org>.

Mentoring in childcare has been documented in the literature for the past 10–15 years (Breunig & Bellm, 1996; Fenichel, 1992). It has been demonstrated to be an effective mode of training/technical assistance (Breunig & Bellm, 1996). However, in the majority of studies conducted there are few, if any, demonstrations that utilize a randomized trial design (Breunig & Bellm, 1996). Many studies track the progress of the intervention group, some studies have comparison groups, but few, if any, have employed a randomized design. This research paper will describe the pre- and post-test data collected as part of a study that has employed a randomized design.

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

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

Methods

Study Design

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

This study employed a randomized design in which a self-selected group of programs and caregivers were randomly assigned to two groups, either the mentoring group or the comparison non-mentoring

comparison/control group. Intervention model mentoring group received intensive mentoring from a seasoned early childhood professional (minimum of 5–7 years of experience in the early childhood field as both a director and teacher) from September to December 2000. The mentoring model consisted of a problem solving approach in which the mentor spent a good deal of time observing in the beginning weeks in order to develop a trusting relationship with the protégé. Once both the mentor and protégé felt comfortable then suggestions could be entertained by the mentor.

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

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

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

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

Table 1
ITERS, Arnett, KIDI, Bloom Scale Scores

All Programs (n = 38)	Pre-Test	Post-Test	Change	Significance
ITERS	134	140	+6	ns
Arnett	30	40	+10	ns
KIDI	14	14	-0-	ns
Bloom	78	79	+1	ns

years of age. They are predominantly Caucasian (81%). Eight percent have associate degrees, 78% have bachelor's degrees, and 14% have master's degrees. They had been employed as directors in their program for an average of 31 months with a range from 1 month to 120 months. Their average pay is between \$20000–25000 per year. Sixty percent have health insurance and 45% have some form of dental or life insurance. Forty-five percent are in a retirement system.

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

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

Results

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

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

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

Although the above results did not reach statistical significance,

Table 2
ITERS

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	134	141	+7	ns
Control Group	137	132	-5	ns

Table 3
Arnett

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	29	40	+11	ns
Control Group	33	40	+7	ns

Table 4
KIDI

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	14	14	-0-	ns
Control Group	14	15	+1	ns

Table 5
Bloom

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	73	74	+1	ns
Control Group	87	91	+4	ns

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

Table 6
Mentoring Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	36	41	.005
Listening activities	8	9	ns
Learning activities	28	31	.05
Interactions	13	13	ns
Adult needs	17	19	ns
Arnett subscales			
Sensitivity	26	31	.001
Appropriate discipline	7	9	.05

Table 7
Control Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	41	42	ns
Listening activities	9	8	ns
Learning activities	29	31	ns
Interactions	15	13	.02
Adult needs	17	17	ns
Arnett subscales			
Sensitivity	28	31	ns
Appropriate discipline	6	7	ns

Discussion

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

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

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

solving activities to improve the overall quality of the interactions and environment of the child care program.

References

- Behar, L. & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, 10, 601–610.
- Breunig, R. & Bellm, D. (1996). *Early childhood mentoring programs: A survey of community initiatives*. Washington, D.C.: Center for the Child Care Workforce.
- Clarke-Stewart, A. (1987). In search of consistencies in child care research. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 105–119). Washington, DC: National Association for the Education of Young Children.
- Feagans, L. & Farran, D. (1979). *Adaptive Language Inventory*. Unpublished instrument. University of North Carolina, Chapel Hill.
- Fenichel, E. (1992). *Learning through supervision and mentorship*. Washington, D.C.: Zero to Three: National Center for Infants, Toddlers and Families.
- Fiene, R. (1995). Utilizing a statewide training system to improve childcare quality. *Child Welfare*, 74(6): 1189–1201.
- Fiene, R. (1996). Using a statistical-indicator methodology for accreditation. In S. Bredekamp (Ed.), *A Decade of Accreditation*. Washington, DC: National Association for the Education of Young Children.
- Galinsky, E., Howes, C., Kontos, S., & Shinn, A. (1994). *The study of children in family childcare and relative care: Highlights of findings*. New York, NY: Families and Work Institute.
- Goelman, H. & Pence, A. (1987). Effects of childcare, family, and individual characteristics on children's language development: The Victoria day care research project. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 89–104). Washington, DC: National Association for the Education of Young Children.
- Griffin, A. & Fiene, R. (1995). *Childcare planning and evaluation*. Washington, DC: Zero to Three, National Center for Infants and Toddlers.
- Harms, T. & Clifford, R. (1980). *The early childhood environment rating scale*. New York: Columbia University Teacher's College Press.
- Helburn, S. (1995). *Cost, quality and child outcomes in child care centers*. Denver, CO: Center for Research in Economics and Social Policy, Department of Economics, University of Colorado.
- Howes, C. (1987). Quality indicators in infant and toddler childcare: The Los Angeles study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 81–88). Washington, DC: National Association for the Education of Young Children.
- Hresko, W., Reid, D. K., & Hammill, P. D. (1981). *The Test of Early Language Development*. Austin, TX: Pro-Ed.
- Iutovich, J., Fiene, R., Johnson, J., Koppel, R., & Langan, F. (1997). *Investing in our children's future*. Erie, PA: Keystone University Research Corporation.
- Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.
- Jorde-Bloom, P. (1988). Assess the climate of your center: Use the early childhood work environment survey. *Day Care and Early Education*, Summer 1988: 9–11.
- Katz, L. (1994). Perspectives on the quality of early childhood programs. *Phi Delta Kappan*.
- Kontos, S. & Fiene, R. (1987). Childcare quality, compliance with regulations, and children's development: The Pennsylvania study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 57–79). Washington, DC: National Association for the Education of Young Children.
- Love, J., Schochet, S., & Meckstroth, R. (1986). *Are they in any real danger: What*

- research does—and doesn't—tell us about childcare quality and children's well being.* Princeton, NJ: Mathematica Policy Research, Inc.
- Maternal and Child Health Bureau (1997). *Stepping stones to Caring for Our Children*, Washington, D.C.: Maternal and Child Health Bureau, Department of Health and Human Services.
- Morgan, G., Costley, C., Genser, A., Goodman, M., Lombardi, J., & McGimsey, J. (1993). *Making a career of it: The states of the states report on career development in early care and education*. Boston: The Center for Career Development in Early Care and Education, Wheelock College.
- NAEYC (1987). *Accreditation standards*, Washington, D.C.: National Association for the Education of Young Children.
- Scarr, S., Eisenberg, E., & Deater-Deckard, R. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9: 131–151.



Child Care Health Consultation Improves Infant and Toddler Care

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

ABSTRACT

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

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

Results: Working with a CCHC effectively improved compliance with some selected health and safety standards. *J Pediatr Health Care.* (2017) 31, 684-694.

KEY WORDS

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

INTRODUCTION

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

In 2013, the Early Childhood Education Linkage System (ECELS), a program of the Pennsylvania (PA) Chapter of the American Academy of Pediatrics (AAP)

Rosemary Johnston, Infant Toddler Quality Improvement Project Coordinator, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Beth A. DelConte, Pediatric Advisor, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Libby Ungvary, Director, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Richard Fiene, Research Psychologist, Research Institute for Key Indicators, Middletown, PA, and Affiliate Professor, Prevention Research Center, The Pennsylvania State University, University Park, PA.

Susan S. Aronson, Pediatric Advisor, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA, and Retired Clinical Professor of Pediatrics, The University of Pennsylvania and The Children's Hospital of Philadelphia, Philadelphia, PA.

This project was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and

Human Services (HHS) under grant No. H25MC26235 Community-Based Integrated Service Systems. The information, content, and/or conclusions in this paper are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, or the U.S. Government.

Conflicts of interest: None to report.

Correspondence: Rosemary Johnston, RN, BSN, MSN, Infant Toddler Quality Improvement Project Coordinator, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System (ECELS), 1400 North Providence Rd., Ste. 3007, Rose Tree Corporate Center II, Media, PA 19063; e-mail: rjrmj@msn.com.

0891-5245/\$36.00

Copyright © 2017 by the National Association of Pediatric Nurse Practitioners. Published by Elsevier Inc. All rights reserved.

Published online August 8, 2017.

<http://dx.doi.org/10.1016/j.pedhc.2017.05.005>

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

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

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

Published studies document the following specific improvements associated with involvement of a CCHC. Sanitation and hygiene reduced respiratory and gastrointestinal illness and days absent for illness among young children in group care (Carabin et al.,

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

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

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

METHODS

Design

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

Selection of the *CFOC3* standards addressed in ITQIP

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

Evaluation plan

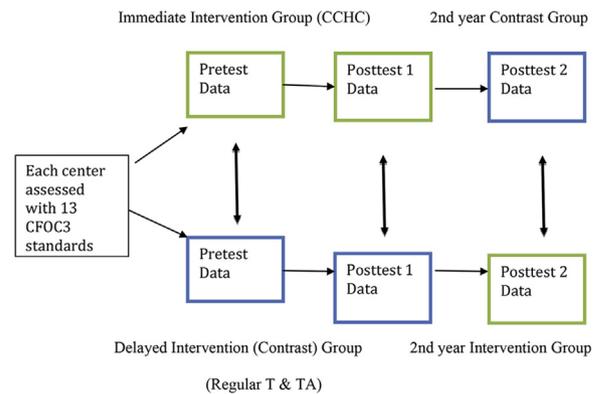
The evaluation plan is a classic randomly assigned crossover clinical trial. See Figure 1 for the evaluation plan logic model.

The ITQIP staff and consultants developed the evaluation tool described below. The ITQIP Project Coordinator (first author) and the evaluators collected data from participating centers at three points: when centers enrolled in the study (Pretest) and then 1 year (Posttest 1) and 2 years later (Posttest 2). One of the consultants (fourth author) compared the two groups on the pretest for equivalency and then on each of the two posttests. These data are discussed in the Results: Immediate Intervention Versus Delayed Intervention (Contrast) Group section. One

BOX 1. *CFOC3* standards chosen for ITQIP

- 1.4.5.2 Child Abuse and Neglect Education
 - 3.4.4.1 Recognizing and Reporting Suspected Child Abuse, Neglect, and Exploitation
 - 2.1.2.1 Personal Caregiver/Teacher Relationships for Infants and Toddlers
 - 2.2.0.2 Limiting Infant/Toddler Time in Crib, High Chair, Car Seat, and other restraining equipment
 - 3.1.3.1 Active Opportunities for Physical Activity
 - 3.1.4.1 Safe Sleep Practices and SIDS Risk Reduction
 - 3.2.1.4 Diaper Changing Procedure
 - 3.2.2.1 Situations That Require Hand Hygiene
 - 3.2.2.2 Handwashing Procedure
 - 3.6.3.3 Training of Caregivers/Teachers to Administer Medication
 - 3.5.0.1 Care Plan for Children with Special Health Care Needs
 - 5.4.5.2 Cribs
 - 7.2.0.1 Immunization Documentation
- Note. *CFOC3*, *Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.)*; *ITQIP*, *Infant-Toddler Quality Improvement Project*; *SIDS*, *sudden infant death syndrome*.

FIGURE 1. Evaluation plan logic model. CCHC, child health care consultant; T, training; TA, technical assistance.



This figure appears in color online at www.jp pedhc.org.

year after the pretest data were collected, the participating centers were switched to a crossover comparison format. At this point, ITQIP ended the subsidy for the CCHCs who were working with the centers in the immediate intervention group and provided the subsidized CCHC linkage to the centers in the delayed intervention (contrast) group.

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

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

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

The ITQIP coordinator scored the evaluator's observations of diapering, hand hygiene, and medication administration. She promptly prepared a summary of all the findings for the center and sent the summary to the center director and the linked CCHC before the first CCHC site visit. The summary delineated areas of strengths and areas to improve based on the evaluation tool results. To facilitate use of the data by the center staff and CCHCs, the summary included the text of the evaluation tool item, the center's score on the item, and the reason why the center met or did not meet the standard. The CCHC contacted the center within 2 weeks after receiving the summary to set up the initial site visit.

Evaluation Tool

The ITQIP staff prepared the items on the evaluation tool from performance guidelines specified in the 13 selected *CFOC3* standards (AAP et al., 2011). ITQIP consultants (fourth and fifth authors) and the ECELS staff reviewed the tool for clarity and validity of content. After several rounds of revisions, the ITQIP coordinator and a prospective ITQIP evaluator field-tested the tool, further revised it, and then field-tested it again, this time

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

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

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

testing for interrater reliability with two evaluators independently and simultaneously using the tool.

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

ITQIP assigned each scorable item to one of the 10 topic areas addressed by the 13 *CFOC3* standards selected for the project (AAP et al., 2011). See Table 1.

Sampling design: Recruitment, random assignment, and retention of centers

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

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

BOX 3. Criteria for scores assigned to items on the evaluation tool

- 0 = Never meets item
- 1 = Partly (<50%) meets item
- 2 = Mostly (≥50%) meets item
- 3 = Fully (100%) meets the item
- NA = Not Applicable
- NOp = Not observed or no opportunity to obtain data
- DK = Don't know (interviewee response)

TABLE 1. Topic areas and number of items to score per topic

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

Note. SIDS, sudden infant death syndrome.
^aSee the narrative for an item-by-item explanation of those items with significance levels (p values) based on the t tests performed on each item.

Centers enrolled in ITQIP agreed to

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

Recruitment and roles of evaluators and CCHCs

Evaluators.

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

The evaluators gave their completed evaluation tools to the ITQIP coordinator to score and summarize. The coordinator reviewed each submitted evalu-

ation tool and then discussed the documentation with the evaluator by phone to make sure the scoring was as intended.

Child Care Health Consultants.

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

During the site visit, the CCHC compared her observations with those in the summary and solicited concerns about health and safety practices from the center's staff. Then the director, program staff, and CCHC chose three of the 10 topics as the primary focus of the center's improvement. The CCHC helped the center staff prepare an action plan to work on the three topic areas they chose.

Action plans included filling gaps in knowledge, developing policies for staff and family handbooks, and improving staff practices. The CCHCs and center directors arranged all subsequent contacts and visits over the next 12 months.

The CCHC helped the center staff prepare an action plan to work on the three topic areas they chose.

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

RESULTS

Descriptive Report

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

59 directors, 348 I/T teachers and 1,490 infants and toddlers were directly involved in ITQIP. Three centers from each group dropped out, leaving 13 centers in each group at the completion of the project (Table 2).

Over the 1-year period of CCHC linkage, 12 of the 32 programs had turnovers of two to four directors. This change in center leadership made the CCHCs' work to improve I/T care very difficult. For the immediate intervention group, three of the original 16 centers withdrew from the project. One center in the delayed intervention (contrast) group closed during the project period; two others withdrew from ITQIP. Some centers dropped out because they were so overwhelmed with maintaining ratios in classrooms and staffing issues that they believed they could not focus on their action plans.

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

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

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

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

The scores used in the quantitative comparisons are the sum of all scores on the Evaluation Tool, not only those

for the topics that the center chose for special focus (Table 4).

Immediate intervention group

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

Delayed intervention (contrast) group

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

Immediate Intervention Versus Delayed Intervention (Contrast) Groups

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

TABLE 2. Location and retention of recruited centers

Region of Pennsylvania	Immediate intervention group			Delayed intervention group		
	Centers recruited	Centers dropped out	Centers completed	Centers recruited	Centers dropped out	Centers completed
Southwest Region (Pittsburgh metropolitan area)	1	0	1	3	1	2
South Central Region (Harrisburg metropolitan area)	4	1	3	2	1	1
Northeast Region (Allentown/Bethlehem/Scranton)	3	0	3	4	0	4
Southeast Region (Philadelphia metropolitan area)	8	2	6	7	1	6
Total	16	3	13	16	3	13

TABLE 3. CFOC3 topics chosen by centers by intervention group

CFOC3 topics	Number of centers in immediate intervention group that chose each topic	Number of centers in delayed intervention (contrast) group that chose each topic
	Safe Sleep Practice	11
Medication Administration	10	6
Child Abuse Prevention	6	1
Care Plans for Children with Special Needs	5	8
Diaper Changing Procedure	4	4
Limited Physical Activity of Infants	2	1
Hand Hygiene	2	5
Immunization	1	0
Personal Relationships	0	1
Active Opportunity for Physical Activity	0	4

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

The crossover comparison results (Figure 2) show the relationship between the immediate intervention and the delayed intervention (contrast) groups in a crossover design. It clearly shows how effective the intervention (pretest to Posttest 1) was for the immediate intervention group and that the effects persisted after 1 year without a subsidized CCHC linkage (Posttest 1 to Posttest 2). It also shows that the intervention was effective when the delayed intervention (contrast) group was switched to receive the CCHC intervention with targeted training, technical assistance, and collaborative consultation a year after their pretest assessment (Posttest 1 to Posttest 2).

For the Immediate Intervention Group After 1 Year of Linkage With a CCHC

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

Medication administration

The director had documentation that the staff who are authorized to give medications have received medica-

tion administration training within the year from a health professional ($p < .001$).

Safe sleep

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

Child abuse

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

Active opportunities for physical activity

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

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

	Intervention group				Delayed intervention (contrast) group			
	Range	Average	%	Possible total	Range	Average	%	Possible total
Pretest	175–267	212 ^a	66	318	164–271	218	68	318
Posttest 1	213–297	254 ^{a,c}	79	318	149–257	221 ^{b,c}	69	318
Posttest 2	137–286	254	79	318	170–283	243 ^b	76	318

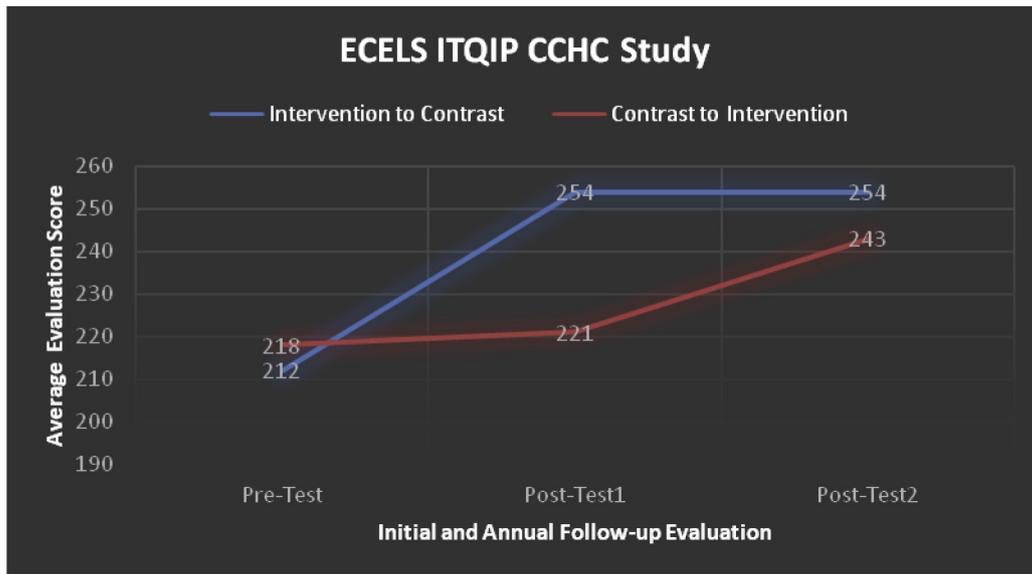
Note. CCHC, child care health consultant.

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

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

^cStatistically significant change ($t = -3.46$; $p < .002$) for Posttest 1 between the immediate intervention group and the delayed intervention (contrast) group.

FIGURE 2. Crossover comparison results. CCHC, child care health consultant; ECELS, Early Childhood Education Linkage System; ITQIP, Infant-Toddler Quality Improvement Project.



went outside except in weather that poses a significant health risk ($p < .05$).

Diaper changing

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

Hand hygiene

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

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

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

Safe sleep

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

every 5 minutes) to be sure that the infant was still breathing ($p < .05$).

Medication administration

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

Diaper changing

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

Special needs

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

Additional Findings of Interest

Immunization documentation

Only one center chose to work on documentation of up-to-date immunization status as an action plan focus. Overall, the immunization data for the two groups showed low compliance with *CFOC3* standard 7.2.0.1 (AAP et al., 2011) and PA's immunization regulations (PA Department of Human Services, 2008). On the pretest, in the immediate intervention centers, 22% of the immunization records for infants and 43% of the immunization records for toddlers were up to date.

Little change occurred for this group on Posttest 1 (36% for infants, 43% for toddlers.) On the pretest for the delayed intervention (contrast) centers, 25% of the immunization records for infants and 40% of the immunizations records for toddlers were up to date. On Posttest 1 the delayed intervention (contrast) centers improved from 25% to 38% for infants but dropped from 40% to 27% of the records for toddlers showing up-to-date vaccines.

Care plans for children with special needs

The data for the two groups showed low compliance with *CFOC3* standard 3.5.0.1 (AAP et al., 2011) that lists the components for care plans. Combining the immediate intervention and delayed intervention (contrast) center findings for this topic, the pretest showed that 66 I/Ts were identified with special health care needs in the 32 centers initially enrolled in ITQIP. Only 15 (23%) of I/Ts with identified special health care needs had any care plan signed by a health care professional. Only 1 of 66 I/Ts with special health care needs had a care plan signed by a health care professional that had all necessary components for optimal daily and/or emergency care. Posttest 2 showed that 39 I/Ts were identified with a special health care need in the remaining 26 centers. For children identified by the centers as having a special health care need, 62% did not have a care plan. Fifteen (38%) of those with identified special health care needs had a care plan signed by a health professional. Four of the 15 care plans had all the required elements. Examples of children who had special needs and had no care plan signed by a health care provider included children with gastroesophageal reflux taking Ranitidine, febrile seizures, asthma, multiple epinephrine autoinjectors onsite, autism, nonfebrile seizures, and torticollis and plagiocephaly, which required that the child wear a helmet each day.

DISCUSSION AND CONCLUSIONS

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

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

and education, medication administration procedure, diaper changing procedures, and care plans for children with special needs with appropriate information and signed by a health care provider.

The data collected by ITQIP show that many children with special needs lacked appropriate care plans. After finding little improvement in the immediate intervention group for centers having care plans with needed elements, ITQIP chose this topic as the focus

of an MCHB-required continuous quality improvement initiative. ITQIP provided an audioconference for the CCHCs and gave them resources for teaching what should be in a care plan. CCHCs reported that they were most successful at helping the centers have complete, useful care plans for children with disease-specific conditions.

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

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

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

The immediate intervention group showed significant improvement in policy development for safe sleep and child abuse and in education about safe sleep practices, preventing child abuse, and medication administration training.

Many of the directors said they appreciated the help they received from the CCHCs that ITQIP linked with their centers.

This selection bias is likely to have influenced the observed improvements.

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

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

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

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

REFERENCES

- Alkon, A., & Bernzweig, J. (2008). *Child care health linkages project evaluation summary*. California Childcare Health Program, UCSF School of Nursing, September 2001 – June 2004. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1525-1446.2008.00689.x/full>
- Alkon, A., Bernzweig, J., Kim, T., Mackie, J., Wolff, M., & Elman, J. (2008). Child care health consultation programs in California: Models, services, and facilitators. *Public Health Nursing, 25*, 126-139.
- Alkon, A., Bernzweig, J., Kim, T., Wolff, M., & Mackie, J. (2009). Child care health consultation improves health and safety policies and practices. *Academic Pediatrics, 9*, 366-370.
- Alkon, A., Crowley, A., Benjamin Neelon, S., Hill, S., Pan, Y., Nguyen, V., ..., Kotch, J. B. (2014). Nutrition and physical activity randomized control trial in child care centers improves knowledge, policies, and children's body mass index. *BMC Public Health, 14*, 215.
- Alkon, A., Nouredini, S., Swartz, A., Sutherland, A. M., Stephens, M., Davidson, N., & Rose, R. (2016). Intergrated Pest Management Intervention in child care centers improves knowledge, pest control, and practices. *Journal of Pediatric Health Care, 30*, e27-e41.
- Alkon, A., Sokal-Gutierrez, K., & Wolff, M. (2002). Child care health consultation improves health knowledge and compliance. *Pediatric Nursing, 28*, 61-65.
- American Academy of Pediatrics, American Public Health Association, National Resource Center for Health and Safety in Child Care and Early Education. (2011). *Caring for our children: National health and safety performance standards; Guidelines for early care and education programs* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics; Washington, DC: American Public Health Association.
- American Academy of Pediatrics, American Public Health Association, National Resource Center for Health and Safety in Child Care and Early Education. 2013. *Stepping stones to caring for our children: National health and safety performance standards; Guidelines for early care and education programs* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics; Washington DC: American Public Health Association.
- Banghart, P., & Kraeder, J. L. (2012). *What can CCDF learn from the research on children's health and safety in child care? Urban Institute Brief #03*. Washington, DC: Urban Institute. Retrieved from <http://www.urban.org/publications/412579.html>
- Carabin, H., Gyorkos, T., Soto, J., Joseph, L., Payment, P., & Collet, J. P. (1999). Effectiveness of a training program in reducing infections in toddlers attending day care centers. *Epidemiology, 10*, 219-227.
- Crowley, A. (2006) *Improving child and family health through child care health consultation. Executive summary*. Beatrice Renfield-Yale School of Nursing Clinical Research Initiatives Fund. Yale University School of Nursing. New Haven, CT: Author.
- Garcia, J. L., Heckman, J. J., Leaf, D. E., & Padros, M. J. (2016). *The life-cycle benefits of an influential early childhood program. Working paper 2016-035*. Chicago, IL: University of Chicago. Retrieved from <https://heckmanequation.org/resource/research-summary-lifecycle-benefits-influential-early-childhood-program/>
- Harms, T., Cryer, D., & Clifford, R. (2006). *Infant/toddler environmental rating scale (revised ed.)*. New York: Teachers College Press.
- Honigfeld, L., Pascoe, T., Macary, S., & Crowley, A. (2017). Promoting children's health in early care and education settings by supporting health consultation. *Impact*. Farmington, CT: Child Health and Development Institute of Connecticut.
- Isbell, P., Kotch, J., Savage, E., Gunn, E., Lu, L., & Weber, D. (2013). Improvement of child care programs' policies, practices, and children's access to health care linked to child care health consultation. *NHSA Dialog: A Research to Practice Journal, 16*, 34-52.
- Kotch, J. B. (2002). *The Quality Enhancement Project for Infants and Toddlers: Executive summary*. Chapel Hill, NC: University of North Carolina at Chapel Hill.
- Kotch, J. B., Isbell, P., Weber, D. J., Nguyen, V., Savage, E., Gunn, E., ..., Allen, J. (2007). Hand-washing and diapering equipment reduces disease among children in out-of-home child care centers. *Pediatrics, 120*, e29-e36.
- Laughlin, L. (2013). Who's minding the kids? Child care arrangements: Spring 2013. Current Population Reports, P70-135. Washington, DC: U.S. Census Bureau. Retrieved from <https://www.census.gov/prod/2013pubs/p70-135.pdf>
- Moon, R., & Oden, R. (2005). Back to sleep: Can we influence child care providers. *Pediatrics, 112*, 878-882.

Organizational Research Services and Geo Education & Research. (2007). Child care health consultation: Evidence based effectiveness. Data from Healthy Child Care Washington evaluation report 2003–2007. Seattle, WA: Washington State Department of Health. Retrieved from https://www.napnap.org/sites/default/files/userfiles/membership/CCS_SIG_Evidence_%20Based_%20CCHP.pdf

Pacific Research and Evaluation. (2008). *Improving the health and safety of children in Oregon's child care: Implementation and outcomes of Oregon Child Care Health Consultation Program*. Portland, OR: Department of Human Services, Public Health Division, Office of Family Health.

Pacific Research and Evaluation. (2007). *Evaluation of the Child Care Health Consultation Demonstration Program: Phase IV Final Report 2007*. Portland, OR: Department of Human Services, Office of Family Health.

Pennsylvania Department of Human Services. (2008). 55 Pa Code, Chapter 3270, Child Care Regulations; 3270.131(a)(5). Retrieved from <http://www.dhs.pa.gov/provider/earlylearning/childcareregulations/>

Pennsylvania Key Program Quality Assessment Team. (2016). [2015–2016 Environmental rating scale reports: ECERS-R, ECERS-3, ITERS-R]. Unpublished raw data.

Pennsylvania Office of Child Development and Early Learning. (2010). *Demonstrating quality: Pennsylvania Keystone STARS 2010*

Program report. Harrisburg, PA: Author. Retrieved from <http://www.ocdelresearch.org/Reports/Forms/AllItems.aspx?RootFolder=%2FReports%2FKeystone%20STARS&FolderCTID=0x01200092EA27E29EEE3E4AAE2D4C5508AC9E5A&View={5EEC6855-F8A8-486E-B6E0-FE6B9FDEBE2E}>

Ramler, M., Nakatsukasa-Ono, W., Loe, C., & Harris, K. (2006). *The influence of child care health consultants in promoting children's health and well-being: A report on selected resources*. Rockville, MD: Maternal Child Health Bureau.

Roberts, L., Smith, W., Jorm, L., Patel, M., Douglas, R. M., & McGilchrist, C. (2000a). Effect of infection control measures on the frequency of upper respiratory infection in child care: A randomized, controlled study. *Pediatrics*, 105, 738-742.

Roberts, L., Smith, W., Jorm, L., Patel, M., Douglas, R. M., & McGilchrist, C. (2000b). Effect of infection control measures on the frequency of diarrheal episodes in child care: A randomized, controlled study. *Pediatrics*, 105, 743-746.

Weinburg, S. (2017). WellCareTracker™ [Computer software]. Retrieved from www.wellcaretracker.org

Zaslow, M., Tout, K., & Halle, T. (2012). *On-site approaches to quality improvement in quality rating and improvement systems: Building on the research on coaching. Research-to-Policy, Research-to-Practice Brief OPRE 2012-40*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

ORDER TODAY

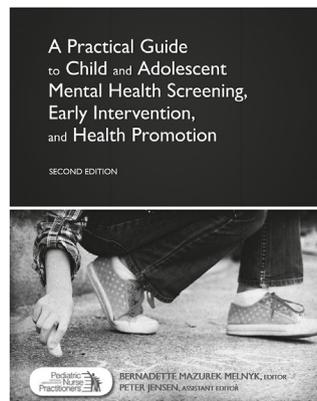
Now Available!

NAPNAP's Practical Guide to Child and Adolescent Mental Health Screening, Early Intervention, and Health Promotion, 2nd Edition

Edited by **Bernadette Mazurek Melnyk**, PhD, RN, CPNP/PMHNP, FAANP, FNAP, FAAN and assistant editor, **Peter Jensen**, MD, expert practitioners in child and adolescent mental health

This practical guide is designed for pediatric healthcare providers including PNP's, FNP's, pediatricians, family physicians and licensed counselors and includes:

- Information focused on the early identification and prevention of childhood and adolescent mental health issues
- Easy access to the important facts about each mental health condition and its corresponding Diagnostic and Statistical Manual of Mental Disorders (DSM-5) criteria
- Valuable screening tools, web-based resources, and educational handouts for parents, school-age children, and teens



Visit www.napnap.org to order your copy of the guide today!

- Maslow, A. H. (1955). Deficiency motivation and growth motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln, NE: University of Nebraska Press.
- National Association of State Boards of Education. (1988). *Right from the start* (report of the NASBE task force on early childhood education). Alexandria, VA: Author.
- National Association of State Boards of Education. (1992). *Caring communities: Supporting young children and families* (report of the national task force on school readiness) (p. 10). Alexandria, VA: Author.
- National Association for the Education of Young Children. (1991). *Accreditation criteria & procedures of the National Academy of Early Childhood Programs*. Washington, DC: Author.
- Pelligrini, A. D., & Glickman, C. D. (1991). Measuring kindergartners' social competence. In *ERIC Digest*. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.
- Pelligrini, A. (1980). The relationship between kindergartens' play and achievement in prereading, language, and writing. *Psychology in the Schools*, 17, 530–535.
- Piaget, J. (1952). *The origins of intelligence*. New York: International Universities Press.
- Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27*. Ypsilanti, MI: High/Scope Press.
- Sibley, A., & Shim, M. A. (1992). *Assessment profile for family day care study guide*. Atlanta, GA: Quality Assist, Inc.
- Smilansky, S., & Shefatya, L. (1990). *Facilitating play: A medium for promoting cognitive, socioemotional, and academic development in young children*. Gaithersburg, MD: Psychosocial and Educational Publications.
- U.S. Department of Education. (1991). *America 2000: An education strategy*. Washington, DC: Author.
- Zero to Three/National Center for Clinical Infant Programs. (1992). *Heart Start: The emotional foundations of school readiness*. Arlington, VA: Author.

Utilizing a Statewide Training System to Improve Child Day Care Quality

Richard Fiene

This article describes Pennsylvania's comprehensive child day care and early childhood development training system, with particular attention to ECELS—the Early Childhood Education Linkage System of the Pennsylvania chapter of the American Academy of Pediatrics—and its immunization initiative. The initiative was established with the state Bureau of Child Day Care Services to improve the overall immunization status of all children in child day care in Pennsylvania. An early childhood education/child day care quality improvement model is briefly discussed to demonstrate how licensing/monitoring data can be tied to a training and technical assistance program.

Richard Fiene, Ph.D., is Research Psychologist, Bureau of Child Day Care, and Adjunct Professor of Psychology and Education, Penn State, Harrisburg, PA.

With the advent of the federal Child Care and Development Block Grant (CCDBG) in 1991, states have implemented initiatives to promote program quality that range from loans to child day care programs to large-scale training systems.* Allowable projects are in five major areas: resource and referral activities, training, licensing improvement, staff salaries, and child day care loans. Many states have used this funding to establish and implement training systems [Morgan et al. 1993].

This article describes Pennsylvania's development of a comprehensive child care and early childhood development training system and the system's initial results, followed by a discussion of the Early Childhood Education Linkage System (ECELS) of the Pennsylvania chapter of the American Academy of Pediatrics. The Pennsylvania chapter has established an innovative system for monitoring and improving childhood immunization status that can serve as a model of how states can link licensing/monitoring data to technical assistance/training system interventions [Fiene 1992].

In 1992, the Pennsylvania Department of Public Welfare (DPW) instituted several collaborative projects to increase training opportunities for child care providers in the Commonwealth. These initiatives were begun for several reasons: (1) DPW wanted to assure that high-quality training would continue to be offered; (2) periodic, ongoing training (six hours per calendar year) was mandated by DPW for all child care workers; and (3) DPW was able to subsidize the training to make it affordable for more child care providers. Pennsylvania used federal and state funds¹ during state fiscal year 1991–1992 to support the statewide training system. Only Texas spent more CCDBG funds on training [Morgan et al. 1993] than Pennsylvania.

*In 1991, CCDBG set aside 6.25% of \$731,915,000 of federal funds for program quality initiatives.

During 1992–1993, Pennsylvania's training system was expanded and renamed the Child Care and Early Childhood Development (CCECD) training system. Approximately 50,000 training encounters were provided during that period, garnering immediate national attention for Pennsylvania for its training initiatives. ECELS, the Keystone University Research Corporation Home-Based Project, and the Southeast Regional School-Age Training Project were all recognized nationally for their innovation and leadership in the early childhood development and child day care training field [Morgan et al. 1993]. It is these initiatives that are described in this article.

During 1993–1994, the major training initiatives continued, with much fine tuning and expansion in areas such as mentoring projects with local programs, additional articulation agreements with local colleges and universities, experimentation with video training, increased input from the Child Care and Early Childhood Development Training Advisory Committee, shared resources and information with the Alliance of Early Childhood Professional Preparation (the state higher education advisory committee) and the Family-Focused Early Intervention System, and expansion of technical assistance opportunities. In all, approximately 40,000 training encounters were provided by eight training contractors during 1993–1994.

Some training programs were arranged or delivered by Marywood College, Keystone Junior College, and Pennsylvania State University. A school-based program was housed within the Central Intermediate Unit in the state college area. Other training programs were housed with nonprofit or corporate entities, such as Community Services for Children, Keystone University Research Corporation, Pittsburgh YMCA, Montgomery Child Care Association, and the Pennsylvania chapter of the American Academy of Pediatrics. Funding and contracts were awarded on a competitive basis.

Other major initiatives were developed as a result of the Pennsylvania Child Care and Early Childhood Development

(CCECD) training system. An early childhood development curriculum continues to be refined and expanded as courses are offered to graduate level teachers and directors. The Child Development Associate (CDA) scholarship program continues to grow. A new early childhood education loan forgiveness program has been instituted and should attract additional highly qualified staff members to teach in child care programs.

At present, training contractors offer classes and courses to child day care staff members across Pennsylvania in the following seven categories: developmentally appropriate practices; emergent literacy; child development; discipline, growth-producing relationships, and interpersonal skill development; health and safety; program administration; and collaboration with community resources and parents.

A description of each of the initiatives and projects follows, with types of services, targeted populations, and service areas noted. The Early Childhood Education Linkage System (ECELS) is highlighted because its major immunization initiative is of special interest to state agencies.

The CCECD Training System

Child Day Care Centers

Licensed child day care centers employ the greatest number of child day care staff. Two providers of training are responsible for assuring that high-quality training sessions on child development and child care operations are available and accessible to child care center staff members (teachers, aides, directors, etc.). Louise Child Care (the Training Connection) serves the western and central regions of Pennsylvania (47 of its 67 counties); Marywood College serves eastern Pennsylvania (the remaining 20 counties).

From July 1993 to June 1994, these two training contracts served over 25,000 child day care employees. Trainees could choose from several options: general workshops, agency-specific

workshops, subsidized early childhood college courses, on-site mentoring sessions, and subsidized conference attendance.

To provide the many training programs needed for licensed child care teaching staff members and to maintain cost-effectiveness, Marywood College and the Training Connection forged partnerships with other organizations and subcontractors. Using regionalized subcontractors enabled them to hold training sessions simultaneously at several sites. Both Marywood College and the Training Connection provided training outlines to subcontractors and reviewed course materials to ensure the quality and consistency of training. All curriculum content was approved by the State Training Director. As of January 1995, over 400 trainers were in the CCECD training system.

Additionally, the two training providers developed articulation agreements (contractual arrangements with local colleges and universities) with several colleges to offer child day care staff members college credits for attending a specified series of training workshops. Thus, child care staff members could advance on the career ladder while implementing quality programs and practices. Helping staff members obtain training necessary for promotion within the child day care and early childhood education field may reduce staff turnover.

School-Age Child Day Care

School-age child care (SACC) programs were developed to provide children from five to 15 years of age with a safe and developmentally appropriate alternative to nonsupervised care. Child day care for school-age children is provided through a diverse network of organizations that includes churches, synagogues, Boys and Girls Clubs, YM/YWCAs, traditional child day care centers, family child care homes, municipal recreation departments, schools, community organizations, and neighborhood centers.

The Pennsylvania Department of Public Welfare funds four organizations to implement, administer, and assure the quality of

services for latchkey children: Community Services for Children (Northeast Region Training Contractor), Day Care Association of Montgomery County (Southeast Region Training Contractor), YMCA of Pittsburgh (Western Region Training Contractor), and Central Intermediate Unit (Central Region Training Contractor). All offer training and technical assistance for quality improvement of existing SACC programs. The four organizations have conducted 300 training programs, educating 4,466 staff members.

Family Child Care Homes

A significant number of children in Pennsylvania receive child day care services in the homes of family child care providers rather than in child day care centers and nursery schools. These providers run small programs that serve from four to as many as 12 children. Home-based providers may receive training through workshops, a training refund program, or a Child Development Associate (CDA) scholarship assistance program [Fiene 1993].

Keystone University Research Corporation (KURC) has served family child care providers with refund/voucher and CDA scholarship programs for over seven years. The refund/voucher program allows providers to attend preapproved training programs of their choice and submit bills for reasonable costs to KURC for reimbursement. Acceptable training options include conferences, seminars, college courses (reimbursed at \$250 per course), workshops, and on-site training. To receive reimbursement, providers must submit evidence of their family child care home registration (family child care homes) or license (group child day care homes), the training must be "appropriate," and the training must last at least six hours. Appropriate training topics include health matters such as first aid and CPR, child development, business management, working with parents, and developmentally appropriate practices and curriculum. A total of 896 reimbursements were approved in 1993–1994.

For child care providers and staff members with low incomes, KURC offers scholarships and scholarship assistance to

help them obtain a CDA credential. The CDA credential indicates that the provider has completed a prescribed course or series of training sessions related to child day care and early childhood education. During 1993–1994, 176 child care staff members and providers received scholarships or financial assistance.

In addition to these ongoing services, KURC offered 300 direct training workshops. With the addition of more direct trainers, the program was able to serve a total of 4,865 child care employees in 1993–1994. KURC also distributes a newsletter and operates a toll-free telephone hotline to make information accessible to child day care providers. The use of both services continued to increase during 1993–1994.

Many families purchase child day care services from unregulated and/or minimally certified providers. It is difficult to assess, improve, or maintain quality at these sites because they rarely come to the attention of officials. In fact, it is virtually impossible to determine the number of providers who operate this way. During the 1992–1993 fiscal year, the Pennsylvania Department of Public Welfare became partners with the Penn State Cooperative Extension to implement the Better Kid Care program in response to the needs of these parents and providers.

The Better Kid Care program offers educational materials—a kit with information on health and safety, child development, and age-appropriate activities for children, along with six learn-at-home programs called "Caring for Kids." Child care workers or parents can obtain the materials by writing or calling, via toll-free hotline, the Penn State Extension offices, which are located in all 67 counties in Pennsylvania. Parents are encouraged to share information with child care providers who might be reluctant to draw attention to themselves by requesting information. More than 5,000 people were enrolled in the project during 1993–1994.

In addition, the 67 Penn State Cooperative Extension county offices conducted workshops for child care providers. Training in various topics in early childhood development and education was provided to more than 2,400 people.

Third-Party Evaluation

A third-party evaluation [Johnson 1994] is being completed at this writing on the overall CCECD system. It has been apparent that the six hours of inservice training per year are not adequate to produce significant program improvements in the classroom by teachers. Regulatory recommendations from the evaluation call for an increase in the six hours of inservice training to the 24 hours per year recommended by the National Association for the Education of Young Children or—preferably—to the 40 hours per year recommended by the Child Welfare League of America.

The evaluation does indicate improvement in classroom implementation of developmentally appropriate practices on the part of staff members who had 20 or more hours of classes and/or workshops on these subjects. Staff members who had fewer than 20 hours of this kind of training did not demonstrate the same level of change in their classroom implementation skills—a key finding.

The evaluation has also uncovered a need to tie training to ongoing monitoring and licensing. This interface of data and technical assistance, which few states use, can serve as a critical link in improving the quality of all child care programs. ECELS—the Early Childhood Education Linkage System—a national demonstration project, has been successful in providing many of these linkages.

The Early Childhood Education Linkage System

The Early Childhood Education Linkage System (ECELS), funded in part by the Pennsylvania Department of Public Welfare and the Pennsylvania Department of Health, collaborates with public and private agencies to improve the well-being of children in the state's early childhood programs. ECELS services include a registry of professional health consultants for technical assistance and local linkage with child care providers; a toll-free information hotline; a free lending library; a quarterly newsletter;

and an arrangement of linkages among community-based trainers and child day care providers.

The early childhood program quality improvement model (PQIM) drawn from ECELS is a conceptual model for interfacing the child care licensing system with the training and technical assistance system [Fiene 1992]. The licensing system literally drives the training and technical assistance system by the use of information derived from licensing data. Appropriate technical assistance and training modalities can be targeted to major areas of noncompliance with regulations. Interventions can be determined by type of provider (center-based or home-based), by geographic area (urban or rural), by size of program, and so on.

Training opportunities include information on evaluating data from immunization and preventive health service records, preventing traffic injuries, and promoting general health and safety. The American Red Cross Child Care Course and self-learning modules on health and safety for family child care providers are also offered. Child day care providers participated in 7,772 units of ECELS training during 1993–1994.

Other noteworthy accomplishments of ECELS in 1993–94 included responding to over 4,000 requests for health information; conference presentations on infection control, health policies, and illness prevention; publication of the quarterly HEALTH LINK newsletter (circulation 15,000+); and recruitment of health professional consultants to assist child day care programs with specific health-related issues and problems.

The Immunization Initiative

The immunization initiative is the product of a unique relationship between ECELS and Pennsylvania. The Bureau of Child Day Care Services licensing staff regularly collects licensing information from providers of child day care services, including sampling records to check the immunization status of children. Over the past several years, the state licensing office determined, both from a sample of programs studied by ECELS and from the

ongoing efforts of the licensing staff, that children in child day care settings lacked up-to-date immunizations, contrary to the state's child day care regulations. Regulatory compliance was only 80% to 85% statewide, in contrast to a national goal of 95% for group child day care settings. Based upon these results, a partnership was forged between the Pennsylvania Bureau of Child Day Care Services and ECELS to increase compliance regarding immunizations in child day care programs.

On a quarterly basis, data gathered by the licensing staff are aggregated by the research division of the bureau and shared with ECELS for follow-up. ECELS follows up with those programs that have the greatest noncompliance with immunization regulations and provides technical assistance and linkage to pediatric services in the community. The director of ECELS is a past president of the Pennsylvania chapter of the American Academy of Pediatrics. Her success in following up with her peers is much greater than what licensing staff members might achieve with the same pediatricians.

This monitoring system has great potential, especially when the federal government's CCDBG efforts to increase the monitoring of the immunization status of children in federally supported child day care are considered. It is an example of how licensing data can be used to drive the focus of a state's training and technical assistance efforts, a subject that has been reviewed extensively [Fiene & Nixon 1985; Fiene 1988, 1992; Aronson 1994] in the early childhood program quality research literature.

The early childhood program quality improvement model has other components that can be used to improve programs that are not a formal part of the training and technical assistance system. Self-assessment as used in accreditation of early childhood programs provides one example of such an approach; site inspection findings from licensing observations or through Head Start performance reviews is another. This program quality improvement model offers a multidimensional approach to the quality enhancement of early childhood programs. A critical

component of the model is the linking of data from monitoring systems to training and technical assistance systems so that interventions can be customized to the needs of individual programs.

Conclusion

The Pennsylvania Child Care and Early Childhood Development training system has offered all early childhood personnel numerous training opportunities to achieve quality practice. Focus groups have been established to obtain feedback from trainees, providers, agencies, and trainers on improving the existing system. Moving toward a seamless early child care and education professional development system is the goal. Expansion of articulation agreements with local colleges and universities is occurring throughout the state. Additional linkages between the licensing and technical assistance/training systems will be explored.

As is true with any statewide training system, constant improvements are always being considered. The Pennsylvania CCECD training system is no exception. For fiscal year 1995–1996, the training system has been locally based by using the Penn State University Cooperative Extension offices as resource training centers in each county. This administrative change moves Pennsylvania's CCECD training system one step closer to a seamless early child care and education professional development system by establishing one statewide training contractor where previously there were eight.

Finally, ECELS's successful tying of licensing/monitoring data to a technical assistance/training system demonstrates that this approach is both cost efficient and effective for program quality improvement in child day care and early childhood programs. The new National Health and Safety Performance Standards for Out-of-Home Child Care Programs [American Public Health Association/American Academy of Pediatrics 1992] are being given priority by the National Center for the Education in

Maternal and Child Health. With selected high-priority standards, it will be possible to target technical assistance and training to key risk areas. ♦

Note

1. Child Care and Development Block Grant (CCDBG) funds (\$2,105,178) as well as state funds (\$617,607) and other federal funding (Child Development Associate (CDA) Scholarship funds = \$100,000 and federal Dependent Care Block Grant funds = \$366,000) were used to fund the Pennsylvania Child Care/Early Childhood Development (CCECD) training system. The CCDBG block grant is a major child care and early childhood development federal funding stream that provides direct child day care service funds and program quality funds to states. The CDA program offers scholarships to low-income and needy students who are pursuing their CDA credential. The Dependent Care Block Grant is a federal funding stream earmarked for programs for school-age children.

References

- American Public Health Association/American Academy of Pediatrics. (1992). *National health and safety performance standards*. Arlington, VA: National Center for Education in Maternal and Child Health.
- Aronson, S. (1994). The science behind the American Public Health Association/American Academy of Pediatrics national health & safety guidelines for child-care programs. *Pediatrics*, *84*, 1101–1104.
- Fiene, R., & Nixon, M. (1985). Instrument-based program monitoring and the indicator checklist for child care. *Child Care Quarterly*, *14*, 198–214.
- Fiene, R. (1988). Human services instrument-based program monitoring and indicator systems. In B. Glastonburg, W. LaMendola, & S. Toole (Eds.), *Information Technology and the Human Services* (pp. 185–190). Chichester, England: John Wiley and Sons.
- Fiene, R. (1992, June). *Measuring child care quality*. Paper presented at the International Conference on Child Day Care Health: Science, Prevention, and Practice, Atlanta, GA.
- Fiene, R. (1993). *Child care and early childhood development annual report*. Harrisburg, PA: Office of Children, Youth and Families.

- Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.
- Morgan, G., Costley, J. B., Genser, A., Goodman, I. F., Lombardi, J., & McGimsey, B. (1993). *Making a career of it*. Boston: The Center for Career Development in Early Care and Education, Wheelock College.

PROFESSIONAL DEVELOPMENT AND THE QUALITY OF CHILD CARE: AN ASSESSMENT OF PENNSYLVANIA'S CHILD CARE TRAINING SYSTEM

Joyce Iutcovich, Richard Fiene, James Johnson,
Ross Koppell and Frances Langan

ABSTRACT

The education and training of child care workers are viewed as keys to improving classroom/caregiver dynamics and the overall quality of child care. This assessment of the Pennsylvania Child Care/Early Childhood Development Training System offers an analysis of this hypothesis. The research was designed for dual purposes: to identify training needs for Pennsylvania child care providers and to assess the impact of training and work environment on the quality of care. The results highlight specific areas where there are needs for training and reveal a clear association between opportunities for professional growth and the quality of care.

INTRODUCTION

The care of children and concerns about their future are of great importance to our society. The current trend in public opinion and political action highlights our concern about children and their welfare. According to public polls "the fastest growing segment of the electorate is the one concerned about protecting

Early Education and Care, and Reconceptualizing Play, Volume 11, pages 115-168.

Copyright © 2001 by Elsevier Science Ltd.

All rights of reproduction in any form reserved.

ISBN: 0-7623-0810-9

children and helping parents be good parents" (McAllister, 1997, p. 36). Further, we have seen new research on the impact of a child's early experiences on how his or her brain is "wired." In an effort to bring attention to this important new research on brain development and its implications for public policy, the Families and Work Institute initiated the Early Childhood Public Engagement Campaign. A White House Conference on Early Childhood Development and a television special, *I Am Your Child*, launched this campaign in early 1997. Another White House Conference on child care was held in October 1997. Politicians have been quick to notice that children's issues strike a special chord with Americans – hence the plethora of new initiatives aimed at the young.

All of this attention on children's issues is heartening in an era of budget reallocation, welfare reform, and the move to eliminate *Big Government*. However, the extent to which all this *talk* will be translated into *action* is yet to be determined. Regardless, this public attention has brought into focus an area of critical need in our society – quality child care. With the dramatic rise in the number of mothers with small children in the labor force, the need for child care services and the maintenance of quality programs throughout the nation cannot be denied (Katz, 1994). In response to this increased demand there has been a significant rise in the number of licensed child care centers and home-based child care providers – not to mention unregulated child care settings. Welfare reform legislation has also resulted in an increase in mothers needing child care services as they move into the labor force. Some welfare-to-work mothers have been encouraged to provide home-based child care to neighbors and relatives to help meet this increased demand for child care.

Thus, as the need increases and child care facilities spring up to meet the growing demand – both regulated and unregulated – the concern over quality becomes more pressing. A study conducted by Mathematica Policy Research for the U.S. Department of Education (1990) reports that the quality of care will be jeopardized with the trend of serving more children with fewer workers. More recent studies have determined that there is far too little good child care in the United States. Only 14% of center care, 12% of family child care, and an even lower percentage of infant care can be rated as good in this country (Galinsky et al., 1994; Helburn et al., 1995).

Given this state of affairs, research on child care and factors associated with quality care are very important, particularly if they have implications for public policy. State regulations play a key role in ensuring that programs comply with minimum standards regarding structural features and staff qualifications. But minimum standards related to child/staff ratios and educational level of staff are not enough. Other dimensions found to be associated with quality care are

classroom/caregiver dynamics (including caregivers' sensitivity and use of developmentally appropriate practice) and staff characteristics such as specialized education, training, and experience (Love, Schochet & Meckstroth, 1996; Barbour, Peters & Baptiste, 1995).

Education and training of child care workers are viewed as keys to improving classroom/caregiver dynamics and quality of care. But not all education and training are equally effective. The Center for Career Development in Early Care and Education at Wheelock College (newly named as the Wheelock College Institute for Leadership and Career Initiatives) has emphasized the importance of professional development programs for child care providers. The model developed by the Center focuses on linkages between education and training and development of new career opportunities for early childhood practitioners (Morgan et al., 1993). Having all training opportunities build on one another, offering incentives for practitioners to obtain training, and specifying a core body of knowledge for all early childhood care and education practitioners are particularly important elements of a model program for career development. Additionally, the Center posits that a comprehensive, coordinated system of training and education should include the following features: quality control over training content and trainers; a system for assessing training needs and offering training based on those needs; a system to make information about training easily accessible and widely distributed; a vehicle for tracking provider training; a linkage between training and compensation; and an expanded and coordinated plan for funding training – preferably through public/private partnerships.

The Study of Pennsylvania's Child Care/Early Childhood Development Training System

Pennsylvania has recognized the need to offer training opportunities for child care workers as a means to improve the quality of care. Training for various segments of the child care provider population has been available for over ten years. In 1992 a number of separate training programs were integrated into one system – *The Pennsylvania Child Care/Early Childhood Development Training System* (PA CC/ECD). The Pennsylvania Department of Public Welfare (DPW) was instrumental in the development of this training system and has supported the establishment of an affordable and flexible training system that is based on the principles of early childhood education and child development.

Pennsylvania's child care training initiative began in the early 1990s, as did other statewide training systems. States utilized the program quality portion of the Child Care and Development Block Grant (CCDBG)¹ to fund the

development and implementation of such training systems (Fiene, 1995). The PA CC/ECD Training System was implemented in January 1992 after lengthy public hearings regarding the Child Care and Development Block Grant. Child care advocates expressed a definite need for a comprehensive early childhood training system throughout the state. Advocates felt that a comprehensive training system was a cost-effective way to improve the quality of early childhood programs throughout Pennsylvania.

The PA CC/ECD Training System has experienced a number of system changes since 1992 and several evaluations with the presently described study as just one of these. For example, prior to 1992, the only training available to child care providers was through a home-based, voucher training program. This program proved to be very popular with providers because it gave them ultimate flexibility in the selection of training opportunities. As the training system evolved, the home-based voucher program became part of the overall PA CC/ECD Training System by 1995. However, this program provides very little structure related to course sequencing or focus on core competencies for child caregivers.

Four school-age technical assistance and capacity building projects also existed prior to 1992, but their major focus was not on training. After 1992 this changed and their focus turned to training. In 1995 the four school-age training projects were incorporated into the overall PA CC/ECD Training System. By 1995, all training for center-based, home-based, and school-age providers were under the umbrella of the PA CC/ECD Training System.

The Early Childhood Education Linkage System (ECELS), the program responsible for health and safety training and technical assistance to Pennsylvania's child care providers, presented the American Red Cross Child Care Course throughout the state from 1992 until 1995. In 1995 this course was incorporated into the PA CC/ECD Training System. This completed the coordination of all training activities related to early childhood and child care under the umbrella of PA CC/ECD with the exception of Head Start and early intervention training.

Since 1992, over 50,000 early childhood providers have received an average of three hours of training on an annual basis. The training opportunities offered to providers include workshops, seminars, videos, learn-at-home materials, conferences, satellite teleconferences, mentoring, vouchers for college coursework, and a number of other training opportunities outside the PA CC/ECD system. The PA CC/ECD Training System is a diverse system of training modalities and funding mechanisms. Several of the PA CC/ECD Training System components have been recognized as innovative. For example, the home-based voucher program and ECELS were recognized in *Making a Career*

of It, a report by the Center for Career Development in Early Care and Education at Wheelock College (Morgan et al., 1993). However, a concern was expressed that training opportunities, albeit comprehensive, were not coordinated to lead an individual on a career path. Therefore, several research studies have been undertaken to determine the effectiveness of the overall system and its implementation.

A Penn State University evaluation research initiated in 1992 helped to delineate the need for additional training opportunities for staff. The accumulative amount of training taken over three years was the key variable that predicted positive developmentally appropriate changes in the classroom (Johnson, 1994). However, this study left unanswered questions about what other factors and features of training are associated with child care quality.

There were overlapping concerns, although different purposes, for two studies initiated in 1996. Wheelock College (Stoney et al., 1997) conducted one study, an assessment of the various early childhood training systems in Pennsylvania, to determine how to coordinate the existing PA CC/ECD Training System with other training systems in an effort to develop a full-fledged early childhood career development system within Pennsylvania.

The other study initiated in 1996 is the one reported herein. Recognizing the importance of tracking the impact of this training system on the quality of care, this research was designed for dual purposes: to identify the training needs for Pennsylvania child care providers and to assess the impact of training and work environment on the quality of care in child care sites. In addition, the results of this research effort are compared to earlier Pennsylvania studies that examined the quality of child care. Within these overarching research goals, this study examined the specific research questions delineated below.

Research Questions Related to Training Needs

- What are the perceived needs for training? Do various provider groups have different needs (e.g. center teachers, center directors, group providers, and family providers)?
- What are the observed needs for training as indicated through the site observations of quality of care?
- What are the most important factors affecting the selection of training? How does the director impact this?
- How do providers evaluate the training? What are their perceptions regarding appropriateness, usefulness, applicability, and effectiveness of training in achieving learning objectives? What is their level of interest in training? And how do they think it applies to their work?

- What are the barriers to training? Are the barriers different for the various provider groups?

Research Questions Related to Quality of Care

- How has the quality of care in Pennsylvania child care changed over the years?
- What factors are significant in predicting the quality of care as observed in child care classrooms?
- To what extent do staff background characteristics (e.g. current education, educational goals, age, years in field, and salary) impact the classroom's quality of care?
- How are features of a caregiver's training experience related to classroom quality of care? To what extent does the level of training impact quality? What is the impact of the training's perceived appropriateness, usefulness, applicability, and effectiveness in achieving learning objectives?
- To what extent do teachers' perceptions of organizational climate impact the quality of classroom care?
- What is characteristic of the quality of work life in child care centers in terms of organizational climate, summary of worker values, overall commitment, how the environment resembles an ideal, the importance of educational goals and objectives, and the degree of influence of teaching staff?
- To what extent is a center's organizational climate associated with director background characteristics, aggregate teacher characteristics, site turnover, accreditation status, size of site, and average hours of training per site?
- To what extent is a center's overall quality of care associated with director background characteristics, aggregate teacher characteristics, organizational climate, and other site-level features (e.g. size of center, accreditation status, turnover rate, and average hours of training per year)?

This study seeks to answer these specific research questions. The following sections present a review of the literature related to professional development systems and factors associated with the quality of child care; an overview of the conceptual framework and methodology used to guide the study; a summary of the results; and the implications of the findings for public policy.

PROFESSIONAL DEVELOPMENT AND QUALITY OF CARE

Staff development research and studies on factors associated with the quality of child care always share the same long term goal, typically hold the same theoretical orientation, and often have variables in common within their

research designs. Ultimately, the goal is to bring about optimal experiences for children in child care. Descriptive and explanatory knowledge about early childhood inservice education or staff development and about program quality is needed to achieve this aim. Other related goals can be served at the same time when research adds to an understanding of quality experiences for children in child care, the value of inservice training for staff development, and the relationship between the two.

Staff Development Research

Current education literature addresses a number of issues related to the ongoing professional development of teaching staff. One very important issue concerns the application of knowledge or the ability to transfer learning into practice. What are the most effective strategies used to guarantee the transfer of knowledge into practice? Numerous reasons are provided as to why staff who participate in educational programs do or do not apply in practice what is learned through education. The perception of program participants about the value and practicality of program content, the presence or absence of follow-up strategies, and supervisory attitudes toward changes required to apply what has been learned are all critical in the transfer of learning (Caffarella, 1994).

The value and practicality of a program implies that a training curriculum should be problem-centered and site-specific. According to Jorde-Bloom and Sheerer (1992), training programs should address real issues and concerns that participants face in their work setting on a daily basis; staff development efforts should facilitate interaction between colleagues; staff developers should "take into account the distinctly different orientations, needs, and interests" of program participants; and training content should focus on bridging the gap between theoretical ideas and the practical realities of the work setting. Jorde-Bloom (1998) also emphasizes the importance of staff becoming active participants in identifying program strengths and areas in need of improvement.

Discussions about the characteristics of effective staff development programs have resulted in some key themes. Holt-Reynolds (1995) maintains the importance of being aware of the rationales underlying the use of particular teaching practices. Rather than focusing on skill training as so many pre-service and in-service teacher development programs have done, staff development must be aimed at uncovering and dealing with lay beliefs, attitudes, behaviors, and decision-making strategies that teachers bring to the classroom.

VanderVen (1994) suggests a contextual model of professional development that enhances the current linear model, which is structural. The contextual model recognizes that early childhood is age-specific and is integrated across the domains of care, education, and development; the contextual model is generic and calls for situational application of multidisciplinary knowledge. VanderVen believes that professional development programs should facilitate constructivism and articulation of theory into practice. Knowledge is gained by doing, then reflecting and dialoguing about it – a constructivist model for learning (Jones, 1993).

In light of the contextual model, outcomes-based educational programs have also been recommended as more effective than the competency-based training programs of the past. Representing a paradigm shift, outcome-based programs focus on demonstrating application of knowledge in contextual settings and quality performance of integrated tasks. Simply acquiring knowledge and demonstrating competencies in isolation is not authentic and does not address the importance of making connections between pre-service training (development) and practice (performance).

Finally, as all this relates to teachers in child care settings, career mobility and advancement is seen as a *sine qua non* of professional development programs (Morgan, 1994). According to Morgan a professional development training system should include these components:

Make training count: when substantial training of good quality is offered, it should carry college credit or be transformed to college credit that can be applied to certificate or degree programs.

Improve access to credit-bearing training for practitioners who are already employed, particularly people of color and individuals from low-income populations.

Articulate programs: accept the Child Development Associate (CDA) Credential to count toward an associate degree program; and allow the associate degree program to count in full toward a bachelor's degree program (Morgan, 1994, p. 138).

Given this background on staff development, we now raise the fundamental question: What constitutes a high quality professional development program? The National Association for the Education of Young Children (NAEYC) has provided leadership in professional development models for early childhood educators (Bredenkamp & Willer, 1994). One of NAEYC's top priorities is improving professional preparation programs for the diverse staff who care for young children. A current NAEYC initiative, the National Institute for Early Childhood Professional Development, is a system designed to address the complexity of developing staff involved in the care and education of young children, improving the quality and consistency of professional pre-service and in-service programs, and linking them with improvements in practice.

Through the work of NAEYC, Willer (1994, pp. 17–19) has identified these principles from the work of Epstein (1993) and Modigliani (1993) that lay the foundation for effective professional development processes; they include:

- Professional development is an ongoing process.
- Professional development experiences are most effective when grounded in sound theoretical and philosophical base and structured as a coherent and systematic program.
- Professional development experiences are most successful when they respond to individuals' background, experiences, and the current context of their role.
- Effective professional development opportunities are structured to promote clear linkages between theory and practice.
- Providers of effective professional development experiences have an appropriate knowledge and experience base.
- Effective professional development experiences use an active hands-on approach and stress an interactive approach that encourages students to learn from one another.
- Effective professional development experiences contribute to positive self esteem by acknowledging the skills and resources brought to the training process as opposed to creating feelings of self-doubt or inadequacy by immediately calling into question an individual's current practices.
- Effective professional development experiences provide opportunities for application and reflection, and allow for individuals to be observed and receive feedback upon what has been learned.
- Students and professionals should be involved in the planning and design of their professional development program.

This attention to early childhood professional development comes at a critical time. Research on the background and skills of child caregivers paints a bleak picture. There is significant concern that child caregivers lack the skills, knowledge, and education to appropriately address the developmental needs of children. "Six out of seven child care centers provide care that is mediocre to poor. One in eight might actually be jeopardizing children's safety and development" (Children's Defense Fund, 1998).

The 1993 National Child Care Staffing Study cited low wages as one factor that accounts for poor quality care. Low wages make recruitment and retention of qualified personnel difficult. Another reason for the low quality of child care is inadequate staff training. "Staff education and training are among the most critical elements in improving children's experiences in child care" (Children's Defense Fund, 1998, p. 39). Regardless, many states do not require pre-service

training for teachers in licensed or regulated child care centers. Further, a majority of states require only 12 or fewer hours of annual training (Children's Defense Fund, 1994). Research has shown that a threshold for training to show some impact is around 18 hours (Howes, Smith & Galinsky, 1995).

Quality of Child Care Research

Reviews of the research on the factors related to child care quality (Phillips, 1987; Love et al., 1996; Chung & Stoney, 1997) group the studies into several categories. Some studies address global assessments of child care quality while others focus on the structural dimensions of quality or the dynamic measures of classroom quality. For our purposes, studies focusing on global assessments and structural dimensions of quality care are of particular importance.

Research from the late 80s (Phillips, 1987) identifies the following as key indicators of quality child care:

- The program is licensed.
- The child's interaction with the caregiver is frequent, verbal, and educational, rather than custodial and controlling.
- Children are not left to spend their time in aimless play.
- There is an adequate adult-child ratio and reasonable group size.
- The caregiver has a balanced training in child development, some degree of professional experience in child care, and has been in the program for some period of time.

More recent studies (Helburn, 1995; Phillips, Howes & Whitebook, 1992) confirm the importance of these indicators and identify other factors that are important. In addition, the following features of high-quality child care for preschool children include:

- **Space:** the indoor environment is clean, in good repair, and well-ventilated; classroom space is divided into richly equipped activity areas; fenced outdoor play space is equipped with swing, climbing equipment, tricycles, and a sandbox.
- **Children's activities:** most of the time children work individually or in small groups; children select many of their own activities and learn through experiences relevant to their own lives; caregivers facilitate children's involvement, accept individual differences, and adjust expectations to children's developing capacities.
- **Parent-caregiver interaction:** parents are encouraged to observe and participate in the program; caregivers talk frequently with parents about children's behavior and development.

For infants, the following signs of high-quality child care are in addition to the key indicators identified by Phillips (1987):

- Play materials are appropriate for infants and toddlers and stored on low shelves within easy reach.
- Daily schedule includes time for active play, quiet play, naps, snacks, and meals; it is flexible rather than rigid, to meet the needs of individual children; and the atmosphere is warm supportive, and children are never left unsupervised.
- Caregivers respond promptly to infants' and toddlers' distress; hold, talk, sing, and read to them; interact with children in a contingent manner that respects the individual child's interests and tolerance for stimulation.
- Parents are welcome anytime; caregivers talk frequently with parents about children's behavior and development.

In light of this overview, the underlying theme is the consistency in which the above factors, as indicators of quality, appear in the research findings.

Conceptual Framework

As previously indicated, staff development research and studies on the quality of child care share the same long term goal and typically the same theoretical orientation. The present study, with its twofold purpose of investigating the PA CC/ECD Training System with respect to user perceptions and the relationship between training and program quality, intersects with the current research literature. Accordingly, its long range purpose, its conceptual underpinnings, and its choice of variables and measures are consistent with previous work in these two areas.

A socio-ecological or systems theory perspective provides a framework for this study. This perspective emphasizes reciprocal transactions between individuals and their environments. Individuals' constructions (beliefs and attitudes) of their social environments, rather than some notion of objective reality, are central to personal adaptation and behavior (Bronfenbrenner, 1979; Lewin, 1935). Child care and training workshops are dynamic, psychological entities as well as physical ones. Providers' social role behaviors and interpersonal relations relevant to the care of children are associated with the totality of factors that constitute a particular child care site (i.e. overall staff and program characteristics). Likewise, providers' role behaviors and relations within child care (staff-staff, staff-child, staff-parents) that contribute to program quality are assumed to influence and be influenced by the PA CC/ECD Training System.

The selection of variables and measures involved in this study, the rationales for the choices, how the variables are conceptually organized, and how they are consistent with previous research are described in the remaining part of this section. These variables are organized into categories as depicted in Figs A and B relevant to the two major purposes of the present study.

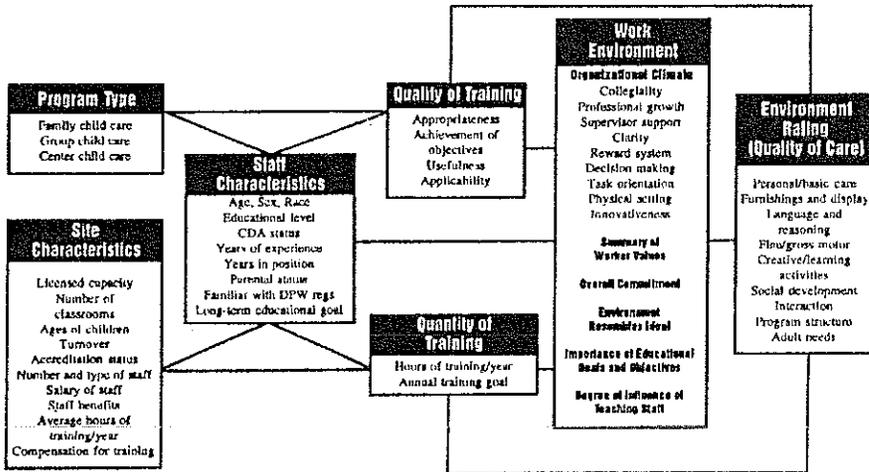


Fig. A. Quality of Care Conceptual Model.

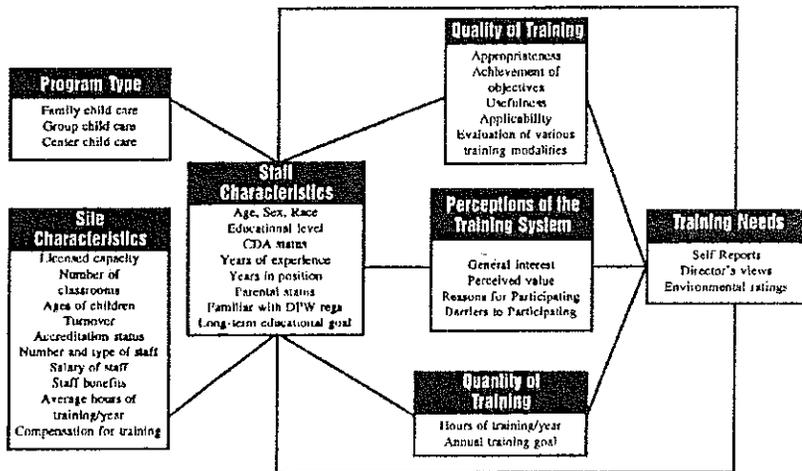


Fig. B. Training System Conceptual Model.

Figure A illustrates how the variables are conceptually organized and associated with levels of child care quality. Quality of child care is operationalized by scores from the Harms and Clifford Environment Rating Scales, while the various dimensions of a child care work environment are measured with Paula Jorde-Bloom's Early Childhood Work Environment Survey. Characteristics of the work environment are viewed as a primary set of intervening variables.

As measures of program quality, three separate environment rating scales were used in this study: the Early Childhood Environment Rating Scale (ECERS), the Infant/Toddler Environment Rating Scale (ITERS), and the Family Day Care Rating Scale (FDCRS). Although each scale has comparable areas that are assessed, the individual items composing each scale do vary depending on the type of child care site or classroom observed. Across each of these scales, the major areas that are assessed relate to furnishings and display; personal/basic care; language and reasoning; fine/gross motor activities; creative/learning activities; social development; interactions; program structure; and adult needs.

As indicators of the various dimensions of an early childhood work environment, the Early Childhood Work Environment Survey (Jorde-Bloom, 1988, 1998) includes a number of distinct conceptual areas. These areas include organizational climate, summary of how current work environment resembles the ideal, importance of educational goals and objectives, and degree of influence of the teaching staff. Organizational climate consists of the collective perceptions of staff regarding the dimensions of collegiality; emphasis on professional growth; degree of supervisor support; clarity of policies and procedures; fairness and equity of the reward system; degree of autonomy in decision making; goal consensus among staff; emphasis on task orientation; extent to which the physical setting facilitates programming; and degree of innovativeness.

Figure A shows an overview of the variables and how they are conceptually organized with respect to the investigation of factors related to the quality of child care. Although the left- to right-hand side ordering of the variable categories in Figure A suggests directionality of effects, it is important to keep in mind that this study is basically descriptive-correlational in nature. The non-experimental, cross-sectional nature of the research design precludes testing directional hypotheses. Program quality could be the cause or the effect of the other variable categories. Nevertheless, the original rationale for selecting this research design centered on the plausible assumption that higher levels of training of personnel in a program would go hand-in-hand with the quality of

care. In addition to organizational climate, certain staff and program characteristics were also expected to show a positive and statistically significant relationship with the quality of care.

The variable categories of program type and program variables shown on the far left-hand side of Figure A are best viewed as moderating variables. These variables suggest data-based comparisons but are not seen as predictors or mediators of quality (with the exception of accreditation status).

Figure B illustrates the relationships among variable categories pertinent to the identification of training needs and user perceptions of the training system. Of major interest, again, are the comparisons involving program type (family child care, group child care, and centers) and type of staff (directors versus teachers). The model included these specific areas: site characteristics, staff characteristics, quality of training, perceptions of the training system, and quantity of training. Training needs and interests were also identified. Questionnaires administered to child care staff were used to identify perceived training needs and interest areas via teacher and director self reports as well as directors' views of staff interests and needs. Needs (as opposed to interests) were also gleaned from information obtained from the environmental rating scales.

METHODOLOGY

As previously indicated, this is a cross-sectional study that collected data from child care sites throughout Pennsylvania. Child care sites were sampled and trained fieldworkers conducted site visits to observe the quality of care in classrooms and to administer questionnaires to child care staff. Specific details about the sampling process, fieldworker training, and data collection instruments are described in the following sections.

Sampling

A stratified systematic sampling process was utilized to identify child care sites for this study. At the time in which we drew the sample, there was a population of approximately 4,144 family child care sites, 590 group homes, and 3,067 child care centers (the *registered* family child care sites and *licensed* group homes and centers). Within each of these separate lists, we then created sampling frames stratified by geographic region. From these stratified lists, we randomly selected a number of child care sites of each type within the various

geographic regions of the state. Our sampling design called for a disproportionate number of sites in each category: 30 family child care homes, 30 group child care homes, and 60 child care centers.

We decided on these numbers for several reasons. First, limited resources and time would not allow us to conduct more than the 120 site visits. Second, to have a sufficient number of group homes to analyze, the number of group child care sites in the sample had to be disproportionate to what they represented in the population. Given the disproportionate nature of the sample, weights were used in any analysis that involved more than one type of child care site.

To encourage voluntary participation in this study, we implemented a number of procedures. First, we initially sent a letter to selected sites to explain the purpose of the study and the importance of the findings for *improving the child care training system in Pennsylvania*. In this letter we explained the advantages of provider participation in the study: receipt of a voucher to purchase children's books/toys from Gryphon House (\$100 for centers and \$50 for family and group homes); an opportunity to have an early childhood professional visit their site and provide some feedback regarding the environment rating scales; and receipt of a certificate acknowledging participation in the study. A follow-up call to the sites was made to further explain the study and encourage their participation. Once a confirmation was received from the site, a fieldworker was assigned to the site to establish a date for a site visit.

A number of the sites initially drawn for the sample were not included in the final total (some were no longer in business, some refused, some could not be visited due to scheduling difficulties). In each case, another randomly drawn site was used as a replacement. Our analysis of the data confirms that the resulting sample was not biased as a result of this replacement; the indicators of quality vary in the expected manner and other site level characteristics reflect known data. The final sample size consisted of 29 family child care homes, 30 group homes, and 60 centers.²

Fieldworker Identification and Training

The importance of having trained observers in a study of this nature cannot be underestimated. For this reason, we took care to identify fieldworkers who were familiar with the Harms and Clifford Environment Rating Scales or with the validation procedures used by the National Association for the Education of Young Children when conducting accreditation site visits. Once fieldworkers were identified, they were sent the training materials (video and manual) for the

Harms and Clifford Environment Rating Scales. Subsequently, a training session was held to review these materials and other procedures to be used in setting up and conducting the site visits. A fieldworker manual was prepared and distributed to everyone; monitoring of their work and progress was conducted from the research office; and inter-rater reliability was determined for a small percentage of each fieldworker's observations for the environment rating scales. The high inter-rater reliability scores indicate consistent use of the scales.³ Furthermore, the overall quality of the data gathered by the fieldworkers attests to their ability.

Data Collection Instruments

The operationalization and measurement of two key areas in this study were previously discussed. Quality of child care was measured through the three Harms and Clifford Environment Rating Scales: ECERS, ITERS, and FDCRS. The scale value for each of the items assessed on these instruments ranges between 1 and 7, where 1 = inadequate, 3 = minimal, 5 = good, and 7 = excellent.⁴

The dimensions of the child care work environment were measured with Paula-Jorde-Bloom's Early Childhood Work Environment Survey (ECWES). There are six separate conceptual areas assessed through this instrument, as identified earlier. For each of the *organizational climate* dimensions, a score of 0 to 10 is calculated by averaging the staff responses to 10 items for each dimension. The *summary of worker values* is indicated by the percentage of staff (0 to 100%) that identify an organizational climate dimension as one of the three most important aspects of their work. *Overall commitment* has a range of values between 0 and 10 where 0 = not committed and 10 = highly committed. Staff perceptions of *how their current work environment compares with their ideal* ranges between 1 = not like my ideal and 5 = like my ideal. The *importance of educational goals and objectives* is indicated by a priority ranking, ranging from 1 = low priority to 6 = high priority. Finally, the *degree of influence of teaching staff regarding organizational decisions* is assessed on a scale of 0 to 10 where 0 = very little influence and 10 = considerable influence.⁵

In addition to these standardized instruments, we developed a series of questions to gather background and training information from both directors and teachers within the child care sites. Although the questions were comparable for directors and teachers and across the type of sites, there were some items that applied only to one or the other. Given this, separate instruments were developed. One instrument was for family providers; one for

directors of small sites (group homes and some small centers); one for directors of centers; one for teachers of small sites; and one for teachers from centers. In the end, we analyzed the data in terms of the type of site (family, group, or center) as well as type of respondent (director or teacher).

The comprehensive background information gathered with these questionnaires included:

- **Director and/or Teacher Background:** age, sex, race, education, years in early childhood field, years with current employer, employment status, salary, long-term educational goal, CDA status, and parental status.
- **Training Background and Assessment:** number of training hours in past three years, annual training goal, evaluation of training system (appropriateness, achievement of goals/objectives, usefulness, applicability), helpfulness of additional training, specialized training, assessment of specific training modalities, decisions about staff training, presence of staff development plans, compensation for training, factors affecting the selection of training, barriers to training, interest in training, and need for additional training in selected topic areas.
- **Site Characteristics:** age of children in facility, type of facility, licensed capacity, number of classrooms, change in licensed capacity in past year, number of paid staff, number of new staff in current year, presence of assistant director, and accreditation status.

FINDINGS

The results of this study address a number of specific research questions within the context of identifying training needs and assessing the factors associated with the quality of care. In presenting the results of the data analysis, we first provide an overview of the background data for each of the provider groups, followed by the findings for the specific research questions.

Background Data on Provider Groups and Child Care Facilities

The socio-demographic characteristics of the provider groups, their training background, and various site characteristics are summarized in Tables 1-3 to give a better understanding of the child care providers and facilities included in this study.

As Table 1 shows, the socio-demographic characteristics of this sample are typical of what we find in national statistics. As expected, the vast majority of providers are female. Their average age is between 34.8 and 45.8 with directors slightly older than teachers. A majority of providers are parents (between 59 to

Table 1. Background Characteristics of Provider Groups.*

Characteristic	Center Directors (N = 60)	Center Teachers (N = 561)	Group Directors (N = 30)	Group Teachers (N = 70)	Family Providers (N = 44)
SEX (% female)	98.3	98.0	100	95.7	93.2
AGE (mean)	41.6	34.9	45.8	34.8	38.8
RACE/ETHNICITY					
White	88.1	82.6	76.7	80.0	66.7
Black	8.5	14.4	20.0	17.1	31.0
Other	3.4	3.0	3.3	2.9	2.4
PARENTAL STATUS (% yes)	72.9	59.1	93.3	70.0	90.9
EDUCATION:					
High school	3.3	32.4	33.3	55.7	54.8
Some college	1.7	22.4	40.0	27.1	33.3
Associate degree	13.3	10.8	13.3	2.9	4.8
Bachelors degree	33.3	23.7	6.7	10.0	4.8
Some graduate	30.0	6.4	3.3	4.3	2.4
Masters degree	13.3	3.1	0.0	0.0	0.0
Post masters	3.3	1.1	3.3	0.0	0.0
Doctorate	1.7	0.0	0.0	0.0	0.0
YEARS IN EARLY CHILDHOOD FIELD (mean)	13.7	6.8	13.1	6.5	7.2
YEARS WITH PRESENT EMPLOYER (mean)	8.6	3.7	9.5	4.2	6.5
EMPLOYMENT STATUS:					
Full-time (35+ hrs)	93.3	62.6	93.3	40.6	80.5
Part-time	6.7	37.4	6.7	59.4	19.5
SALARY (approx. average)	\$19,900/yr	\$6.40/hr	\$17,250/yr	\$5.89/hr	\$12,500/yr
BENEFITS					
Pension		18.5		2.1	
Vision		15.5		2.1	
Dental		32.6		2.1	
Health	N.A.**	48.7	N.A.	2.1	N.A.
Life insurance		23.5		0.0	
Paid maternity		3.2		0.0	
Disability		16.3		2.1	
Education reimbursement		25.9		17.0	

* Percentages are reported except where otherwise noted.

** N.A. = Question not asked of this provider group.

Table 2. Training Background of Provider Groups.*

Characteristic	Center Directors (N = 60)	Center Teachers (N = 561)	Group Directors (N = 30)	Group Teachers (N = 70)	Family Providers (N = 44)
LONG TERM EDUCATIONAL GOAL					
GED/High school	0.0	4.3	6.9	7.4	4.8
Non-credit adult education	1.8	5.8	6.9	13.2	9.5
Early childhood certification	0.0	12.2	13.8	16.2	9.5
Associate degree	1.8	6.8	13.8	4.4	16.7
College degree	5.4	15.9	13.8	8.8	14.3
Graduate degree	57.1	17.6	10.3	10.3	4.8
No long term goals	33.9	37.5	34.5	39.7	40.5
SEEKING CDA CERTIFICATE					
Yes*	1.8	16.9	20.8	16.9	25.6
No	85.5	75.8	79.2	74.6	71.8
Already have	12.7	7.3	0.0	8.5	2.6
TRAINING IN PAST 3 YRS (mean hours)	43.1	18.5	40.3	20.5	20.2
ANNUAL TRAINING GOAL					
6 hours	40.7	67.3	31.0	63.9	53.7
12 hours	27.1	19.0	31.0	21.3	14.6
12+ hours	32.2	13.7	37.9	14.8	31.7
PERSONAL CAREER DEVELOPMENT PLAN (% yes)	N.A.**	55.1	N.A.	55.0	71.1

* Percentages are reported except where otherwise noted.

** N.A. = Question not asked of this provider group.

93%) with center teachers least likely to hold this status. Center directors hold the highest levels of education while group teachers and family providers have the lowest levels. The directors for both centers and group facilities have been in the field of early childhood education longer than the other provider groups (on an average of thirteen years for directors in comparison to approximately seven years for child care teachers and family providers). Center teachers have the least amount of time with their current employer when compared with their total number of years in the field. The vast majority (over 93%) of directors for

Table 3. Facility Characteristics.

Characteristic	Centers (N = 60)	Group Homes (N = 30)	Family Homes (N = 29)
Licensed Capacity (mean)	76.23	13.8	6.6
Number of Classrooms (mean)	4.95	2.1	N.A.**
Number of Children Enrolled (mean)	68.73	15.9	7.2
Age of Children (% of facilities with):			
Birth to 12 months	55.0	66.7	52.3
13-24 months	71.7	80.0	68.2
25-36 months	83.3	90.0	72.7
3-5 years	96.7	96.7	88.6
6-8 years	63.3	60.0	43.2
9+ years	48.3	33.3	22.7
Special needs (% yes)	61.7	16.7	11.4
Number of Paid Staff (mean)	10.93	3.6	N.A.
Assistant Director (% yes)	37.3	35.0	N.A.
Turnover Rate	0.22	0.31	N.A.
Accreditation Status (% yes)	26.3	10.0	22.5

** N.A. = Question not asked of this provider group.

both centers and group facilities are full-time, while a majority of group teachers (59.4%) are part-time. Regarding compensation, group teachers are also the lowest paid (approximately \$5.89/hour), while center directors, on the average, earn the highest salaries – just under \$20,000 per year. Benefits are also not prevalent in the field, although center staff are more likely to have some benefits than are home-based providers. Health benefits are the most common, yet less than half (48.7%) of the center teachers report having this benefit.

Table 2 summarizes the responses to questions that are indicators of the extent to which providers are motivated to pursue additional as well as higher levels of education and training. Over one-third of each provider group indicate

that they have no long-term educational goals. However, center directors are more likely to express a desire for higher education, with 57.1% indicating that a graduate degree is a long-term educational goal. As far as other child care training, a substantial percentage of providers do *not* have a Child Development Associate (CDA) certificate, but center directors (12.7%) are more likely to have the CDA than are other provider groups. Furthermore, directors of both centers and group facilities have, on the average, twice the number of training hours than do teacher and family provider groups. Over the past three years, directors averaged over 40 hours of training, while teachers and family providers averaged around 20 hours (just slightly higher than what is required to meet the state regulations of 6 hours per year). The emphasis on only meeting state requirements is further evidenced by the responses from providers when asked to indicate their annual training goal. A majority in each provider group, except group directors, indicates that completing 6 hours is their goal. Both directors of centers and group facilities, as well as family providers show greater interest in education/training beyond the minimum required. The final indicator of a provider's educational interest and motivation is revealed when asked, "Do you have a plan for your individual career development in early childhood care and education?" More than half of teachers and family providers indicate they have a personal career development plan. This appears to be a higher percentage than expected, given the responses to the other questions related to educational interest and motivation. However, this question did not ask if the plan was written and/or formalized; as such, the responses to this question may include individuals who at a minimum have *thought* about their plans for further training and education.

Characteristics of the sample sites are shown in Table 3. On average, centers have a licensed capacity for 76 children, just under five classrooms, and an enrollment of 69 children. While centers have fewer enrolled children than they are licensed for, both group and family homes have more (probably due to school-age children or children who might not be enrolled for full-time child care). As for the age of children served, children age two through five are most likely to be enrolled in child care. Special needs children are most likely served by centers, not group or family homes. Staffing patterns are also consistent with common knowledge – centers average just under eleven paid staff, while group homes average just fewer than four. Approximately one-third of both centers and group homes have an assistant director. The turnover rate, indicated by the ratio of new staff to total number employed, is slightly higher for group child care (0.31) than it is for centers (0.22). Centers are most likely to be accredited (26.3%) while group homes are least likely (10%).

Training Needs and Perceptions of Current Training System

Perceived Training Needs

The survey instrument distributed to child care staff asked both directors and teachers to identify the need for training in specified training topics.⁶ They were also asked to base their assessment on the *need for training* for child care providers, not just the importance of the topic alone. Table 4 summarizes the responses of these provider groups: center directors, center teachers, group providers,⁷ and family providers.

In examining Table 4, if we rank order the topics in terms of perceived priority, we see that the general topic area of **supervision, motivation, and discipline/guidance of children** is considered an area of very serious need for training. This topic is ranked at the top for all provider groups except family providers who rank it as the second most needed area of training. Family providers identify **fostering social development (e.g. dealing with conflict)** as the top priority for training. These two topics are closely related in that they both deal with the issue of behavior management – a serious concern for providers that is repeatedly expressed by them. A concern over behavior management is further supported by the data. All provider groups rank both topics as either first or second priority for training.

When all topics are listed in rank order (from topics that are a very serious need to topics that are not a priority), there is a high degree of consistency across all provider groups – for center directors and teachers as well as home-based providers. The four areas consistently ranked as priority training topics are:

- supervision, motivation, and discipline/guidance of children
- social development (dealing with conflict)
- child development
- developmentally appropriate practice

In addition, family providers identify **nutrition** and **infant/child development** as important areas of training. Regardless of relative importance and rank order position, providers view none of the training topics specified on the research instrument as *unimportant*. The average scale value for these topics ranged between 1.28 and 2.53—thus, there is no topic area that is viewed as *not a priority* for training.

Training Needs as Observed via the Environment Rating Scales

In addition to the identification of training needs through the self-reports of child care staff, we are able to provide a more objective measure via the Harms

Table 4. Perceived Need for Training in Selected Topic Areas.*

Training Topic	Center Directors (N = 60)	Center Teachers (N = 546)	Group Providers (N = 100)	Family Providers (N = 44)
Child care business, management	2.19 (16)	2.36 (17)	2.43 (19)	2.03 (17)
Child care program development	1.77 (8)	1.87 (9)	2.04 (12)	1.67 (9)
Child development	1.63 (4)	1.57 (3)	1.72 (2)	1.59 (5)
Child/staff health	2.05 (14)	1.77 (5)	1.89 (7)	1.80 (12)
Development appropriate practice	1.43 (3)	1.65 (4)	1.81 (4)	1.66 (8)
Emergent literacy, children's literature or literacy-based socio-dramatic play	1.84 (9)	2.04 (15)	2.05 (13)	1.92 (14)
Emergent numeracy, science for young children	1.74 (6)	1.92 (11)	2.12 (16)	2.00 (16)
Fostering social development (e.g. dealing with conflict)	1.39 (2)	1.56 (2)	1.76 (3)	1.44 (1)
Inclusive/special needs education issues	1.74 (7)	1.78 (6)	2.05 (14)	1.69 (10)
Infant/ Toddler child development/programming	1.88 (10)	1.78 (7)	1.89 (6)	1.51 (3)
Multicultural, gender sensitivity in programming for young children	1.93 (11)	2.02 (14)	2.01 (10)	1.95 (15)
Music, dance, movement for young children	1.98 (13)	1.93 (12)	2.02 (11)	1.89 (13)
Nutrition	2.27 (17)	1.99 (13)	1.88 (5)	1.57 (4)
Personal care routines (naptime, toileting, grooming)	2.46 (19)	2.12 (16)	2.11 (15)	1.60 (6)
Play	1.97 (12)	1.91 (10)	1.96 (8)	1.71 (11)
Supervision, motivation discipline/guidance of children	1.28 (1)	1.51 (1)	1.55 (1)	1.46 (2)
Working with parents/community services	1.73 (5)	1.85 (8)	2.00 (9)	1.64 (7)
Statewide conference on multiple topics	2.48 (20)	2.53 (20)	2.48 (20)	2.22 (20)
Regional conference on multiple topics	2.41 (18)	2.53 (19)	2.36 (18)	2.18 (19)
Mentoring, multiple topics	2.18 (15)	2.45 (18)	2.30 (17)	2.08 (18)

* Perceived need is indicated by the mean score for the provider group on a scale of 1 = a very serious need, 2 = important but not critical, 3 = more would be helpful, and 4 = not a priority; in addition, a rank order of training needs for each provider group is indicated in parentheses.

and Clifford Environment Rating Scales. By identifying areas where child care sites are weak (e.g. where average scores are less than 5), we can specify needed training topics. Table 5 summarizes the average scores for the individual items included in each of the environment rating scales (FDCRS, ITERS, and ECERS).

In analyzing the set of individual items on the three different Harms and Clifford Environment Rating Scales, we see that there are a number of areas that receive a very low rating – below a scale value of 4.00. Items rated this low indicate areas where special attention should be placed in the design and delivery of training. Across all three scales – FDCRS, ITERS, and ECERS – these items are consistently rated low: **cultural awareness, personal grooming, dramatic (pretend) play, and sand and water play**. Furthermore, these areas are rated low in two out of the three environment rating scales: **displays for children** (FDCRS and ITERS), **space alone** (FDCRS and ECERS), **helping infants/toddlers understand language** (FDCRS and ITERS), **art** (ITERS and ECERS), and **blocks** (FDCRS and ITERS).

Overall, the ECERS reveals fewer areas of serious concern (only 16% of the items on this scale have a score below 4.00), while the ITERS reveals the most (46% of the ITERS' items have a score below 4.00). This is consistent with national data on the environment rating scales (Phillips, 1987; Scarr, 1994). Indeed, if we compare the overall average score for each scale, (FDCRS = 4.47; ITERS = 4.26; ECERS = 4.63), the ITERS has the lowest average score. This indicates a need for particular focus on infant/toddler training, a finding that is consistent with anecdotal evidence and comments.

On the other end of the continuum, there are a number of items on each of these scales that score above 5.00, indicating an assessment in the *good* range. Keeping in mind that there are not comparable items across all three scales,⁸ we consistently see these areas rated highly: **nap/rest time, discipline/supervision, provision for parents, informal use of language with infants/toddlers, and health practice and/or policy**. Consistent with our analysis of the items rated poorly, the ECERS fares the best. It has the highest percentage of items (38%) receiving a score above 5.00 (ITERS only has 26% of the items scoring above 5.00, while FDCRS has 23%). There are several points of interest in our examination of these ratings. First, it is noteworthy that the health area received such a positive evaluation. No doubt, concerns about health and safety are of primary importance to parents as well as officials who regulate child care. Second, the high rating for discipline/supervision is paradoxical given the consistent identification of this area by caregivers as one in which they most need training. This illustrates that caregivers are performing better in this area than they perceive; it also reveals that discipline/supervision

Table 5. Average Score on Individual Environment Rating Scale Items.

Scale Item	FDCRS (N=67)*	ITERS (N=36)	ECERS (N=57)*
Furnishings and Display	4.10	4.09	4.57
Furnishings for routine care	4.87	4.53	5.75
Use of furnishings for learning activities	-	4.44	4.14
Furnishings for relaxation and comfort	4.71	3.69	4.13
Room arrangement	-	3.86	4.74
Child-related display	3.01	3.92	4.08
Indoor space arrangement	4.11	-	-
Active physical play	4.55	-	-
Space to be alone			
a. Infants/toddlers	3.43	-	-
b. 2 years and older	3.76	-	-
Personal/Basic Care	4.84	4.66	4.59
Arriving/departing	6.15	5.56	4.71
Meals/snacks	4.72	3.93	4.40
Nap/rest	5.07	5.10	5.64
Diapering/toileting	4.13	3.62	5.05
Personal grooming	3.78	3.71	3.23
Health practice	5.17	4.21	-
Health policy	-	5.63	-
Safety practice	4.86	5.40	-
Safety policy	-	5.46	-
Language and Reasoning	4.54	4.37	4.82
Informal use of language			
a. Infants/toddlers	5.01	5.00	-
b. 2 years and older	4.90	-	4.89
Helping children understand language			
a. infants/toddlers (books & pictures)	3.47	3.74	-
b. 2 years and older	4.29	-	5.02
Helping children use language	4.45	-	4.99
Helping children reason	4.35	-	4.36
Fine/Gross Motor	N.A.	N.A.	5.11
Fine motor			5.41
Supervision (FM)			5.10
GM space			5.02
GM equipment			4.66
GM time			5.21
Supervision (GM)			5.44

Table 5. Continued.

Scale Item	FDCRS (N = 67)*	ITERS (N = 36)	ECERS (N = 57)*
Creative/Learning Activities	4.12	3.39	4.46
Eye-hand coordination	4.48	4.67	—
Active physical play	—	3.53	—
Art	4.08	3.81	3.81
Music and movement	4.76	4.19	5.20
Sand and water play	2.60	3.07	3.75
Dramatic (pretend) play	3.74	3.07	3.62
Blocks	3.88	3.21	4.44
Use of T.V.	4.19	—	—
Schedule of daily activities	4.59	—	4.93
Supervision of play indoors and outdoors	4.79	—	5.51
Cultural awareness	—	1.75	—
Social Development	4.72	N.A.	4.20
Tone	5.73	—	5.34
Discipline	5.59	—	—
Cultural Awareness	2.85	—	2.96
Space (alone)	—	—	3.60
Free play	—	—	4.53
Group time	—	—	4.33
Exceptional provisions	—	—	4.69
Interaction	N.A.	4.98	N.A.
Peer interaction	—	4.93	—
Adult-child interaction	—	4.99	—
Discipline	—	5.01	—
Program Structure	N.A.	4.53	N.A.
Schedule of daily activities	—	3.75	—
Supervision of daily activities	—	4.71	—
Staff cooperation	—	4.98	—
Provisions for exceptional children	—	5.30	—
Adult Needs	5.17	4.28	4.80
Adult personal needs	—	3.31	4.11
Opportunities for professional growth	4.79	3.57	4.50
Adult meeting area	—	4.94	5.10
Provisions for parents	—	5.35	5.51
Relationships with parents	5.37	—	—
Balancing personal and caregiving responsibilities	5.28	—	—

* This is the weighted N since there were observations made in more than one type of child care (i.e. family, group, or center).

** N.A. = Question not applicable for this environment rating scale. Spaces where there are no applicable scores are indicated by "—".

is perhaps one of the most challenging areas in child care and something for which caregivers think they need constant help and support.

Selection of Training

Providers were asked to indicate the importance of a number of factors in their selection of training.⁹ In Table 6 we see, again, there is a high degree of consistency across all provider groups. Providers indicate that their selection is based primarily on their **interest in a topic** and if a topic **helps in understanding children**. Furthermore, center staff (directors and teachers) identify **opportunities for professional development** as important. All provider groups, except center directors, rank **training that offers practical solutions** within the top five on their list of factors that are important in the

Table 6. Factors Affecting the Selection of Training.*

Selection Factors	Center Directors (N = 60)	Center Teachers (N = 546)	Group Providers (N = 100)	Family Providers (N = 44)
Location/convenience	1.19 (5)	1.27 (5)	1.29 (3)	1.08 (3)
Session length	1.51 (10)	1.70 (12)	1.71 (11)	1.54 (10)
Meet state requirements	1.24 (7)	1.40 (8)	1.33 (5)	1.34 (6)
Quality of previous training	1.41 (8)	1.39 (7)	1.49 (8)	1.49 (8)
Cost of training	1.53 (11)	1.66 (11)	1.50 (9)	1.58 (11)
Scheduled times of training	1.13 (4)	1.31 (6)	1.36 (6)	1.03 (1)
Interest in topic/contents	1.10 (1)	1.15 (2)	1.15 (1)	1.21 (4)
Networking opportunities	1.75 (12)	1.81 (13)	1.82 (13)	1.66 (12)
Training organization	1.76 (13)	1.65 (10)	1.72 (12)	1.69 (13)
The trainer	1.48 (9)	1.51 (9)	1.66 (10)	1.54 (9)
Offers practical solutions	1.19 (6)	1.25 (4)	1.32 (4)	1.24 (5)
Helps understand children	1.12 (3)	1.09 (1)	1.16 (2)	1.08 (2)
Professional development	1.10 (2)	1.22 (3)	1.37 (7)	1.35 (7)
Sent by director	N.A.**	1.91 (14)	1.87 (14)***	N.A.**

* Importance of factors in the selection of training is indicated by the mean score for the provider group on a scale of 1 = a very important, 2 = somewhat important, and 3 = not important. In addition, the rank order of the factors in terms of importance is indicated in parentheses.

** N.A. = Not asked of this provider group.

*** This represents the response from the group teachers only.

selection of training. Center directors mention **scheduled times for training** as important. Similarly, home-based providers mention **scheduled times for training or location/convenience** as important factors affecting their selection of training. These priority rankings are congruent with the role responsibilities of center directors and teachers and home-based providers. Directors are responsible for the scheduling of staff at their child care facility, while home-based providers must participate in training that is offered during nonbusiness hours – hence the importance of when training is scheduled. On the other hand, teachers deal with the day-to-day child care activities for which they want practical guidance.

However, all of the factors that might affect selection of training are considered at least *somewhat important* by child care providers. (Note that none of the factors have a mean score above 2.0.) But, in terms of priority, the factors having the least priority across all provider groups are **networking opportunities, training organization, session length, cost of training, and trainer**. The relative unimportance of the cost of training is to be expected. The Pennsylvania child care training system provides training opportunities at no cost, or for a minimal registration fee, therefore cost is not a critical issue. As for the trainer and training organization, it may be that providers are satisfied with current training organizations and trainers (as expressed elsewhere in these data and also in the participant evaluation forms completed for each training session). These data indicate the trainer or training organization may not be as important as other factors in the selection of training.

What is of interest is the relative unimportance of **networking opportunities**. Anecdotally, we often hear that the opportunity to meet and talk with other child care providers is highly valued. On closer inspection, we see that family providers (the provider group that is most isolated from peers), are more likely to consider networking opportunities as important than are the other provider groups. Half of the family providers indicate that networking is a very important factor in their selection of training, while only around one-third of the other provider groups indicate this.

Center teachers also were asked to indicate the importance of **being sent by the director** in selection of training. In comparison to other factors, **being sent by the director** is relatively unimportant – it is ranked at the bottom. Regardless, approximately one-third of the teachers in centers indicate that **being sent by the director** is a *very important* factor. Ideally, directors of child care centers should be working with staff to establish professional development plans that meet the individual needs of workers. However, this question, as asked, does not identify the reason why a director sends staff to a particular training – i.e. whether the selected training corresponds with professional

development needs of staff or whether training is offered at a convenient time and place.

When directors were asked about how decisions are made regarding staff training, just under half of center directors (46.6%) indicated that they “guide the selection but the staff make the final decision.” Whereas, in the group child care situation, 60% of group directors indicated this.

Having a personal plan for career development is related to this decision-making process and the selection of child care training. Whether or not staff have such plans was assessed by asking directors “What percentage of your child care staff have personal plans for career development in early child care and education?” Center directors, on average, indicate that over half (51.9%) of staff have personal plans. In group child care, directors report that only 24.2% of staff have personal career development plans. A much higher percentage (71.1%) of family providers have a plan for personal development as a child care provider. This question does not ask for specific details, therefore the interpretation of what constitutes a *plan* probably varies considerably.

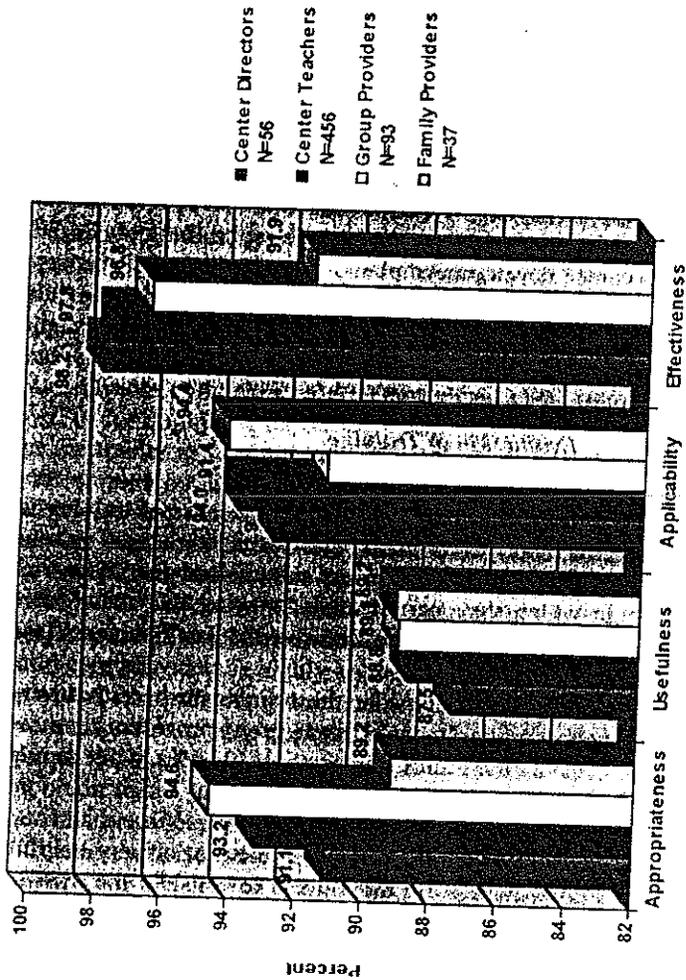
Evaluation of Training

The appropriateness, usefulness, applicability, and effectiveness of training in achieving learning objectives, as perceived by providers, were used as one means to evaluate the training system. Providers were asked for an *overall* assessment of training in which they participated, knowing that many have participated in a number of training opportunities over the past few years (see Fig. 1).

A majority of all provider groups consider the training to be either very appropriate or somewhat appropriate. In comparison, group providers are more likely than the others to consider the training appropriate (94.6%), while family providers are least likely (89.2%).

Providers also positively assess the usefulness of training. More than four-fifths of each provider group consider the training somewhat or very helpful. Comparatively, home-based providers are most likely to consider the training useful (group=89.1% and family=89.2%), while center directors are least likely (87.5%).

The applicability of training (or the knowledge and skills learned) to the work environment should be an important feature of any training system if it is to have an impact. It is impressive that a substantial majority of providers (over 90%) indicate that they could apply *all, a lot, or some* of what they learned in the training to current work.



* For each factor, these data represent the percentage of providers who indicate that:
 • The training was "very appropriate" or "somewhat appropriate" (*Appropriateness*)
 • The training was "very helpful" or "somewhat helpful" in their work (*Usefulness*)
 • They can apply "all," "a lot," or "some" of what they learned in training (*Applicability*)
 • The training goals were "achieved" or "somewhat achieved" (*Effectiveness*)

Fig. 1. Perceived Appropriateness, Usefulness, Applicability, and Effectiveness of Training.*

To assess the perceived effectiveness of training, providers were asked to indicate the extent to which the training goal(s) were achieved, that is, the extent to which they learned the material.¹⁰ As with applicability of training, almost all providers (over 90%) respond that they learned at least some of the material. A slightly smaller percentage of family providers indicate this (91.9% vs. over 96% for the other provider groups).

Overall, the training system is viewed positively by provider groups, as evidenced by response to questions about appropriateness, usefulness, applicability, and effectiveness in achieving learning objectives. The providers consider the training appropriate for their level of knowledge and skill, find it helpful in their current work, indicate they are able to apply what they have learned, and feel training goals have been achieved.

This positive assessment also corresponds with their response when asked about level of interest in training and if more training would be (see Fig. 2).¹¹ As with the other evaluative factors, the level of interest is high among teachers, with over 80% of center and group teachers indicating they are either *interested* or *very interested* in taking training. Furthermore, directors are on target in assessing levels of interest of their staff. As further evidence of the positive evaluation of the training by providers, a substantial percentage (86–100%) indicate that attending more workshops or training will help them in their work.

A final evaluative measure used in assessing the current training system asked about the perceived helpfulness of the various training methods used in the Pennsylvania Child Care/Early Childhood Development Training System. Only directors and family providers were asked about this.¹² Table 7 summarizes the responses for center directors, group directors, and family providers. **On-site training** ranks as the most helpful method by the directors of centers and group homes, while family providers rank it as second most helpful. Center directors and family providers also positively assess **workshops**. While **satellite** and **video** methods of training may be cost effective and efficient in reaching providers in more rural areas, both these methods of training are viewed as less helpful than other methods. Interestingly, both family providers and group directors express a more positive view of these two methods than do center directors.

Barriers to Training

Several factors may limit child care providers from attending training. Providers were asked to indicate the importance of a number of factors that might prevent them from attending training or workshops (see Fig. 3).

Table 7. Perceived Helpfulness of Training Methods.*

Training Method	Center Directors (N = 60)	Group Directors (N = 30)	Family Providers (N = 44)
Workshop	1.26 (2)	1.52 (2)	1.16 (1)
Satellite	2.33 (6)	2.07 (6)	1.84 (6)
Video	1.92 (5)	1.56 (5)	1.60 (4)
On-site Training	1.21 (1)	1.39 (1)	1.42 (2)
Conference	1.57 (4)	1.48 (3)	1.55 (3)
Mentoring	1.38 (3)	1.47(4)	1.78 (5)

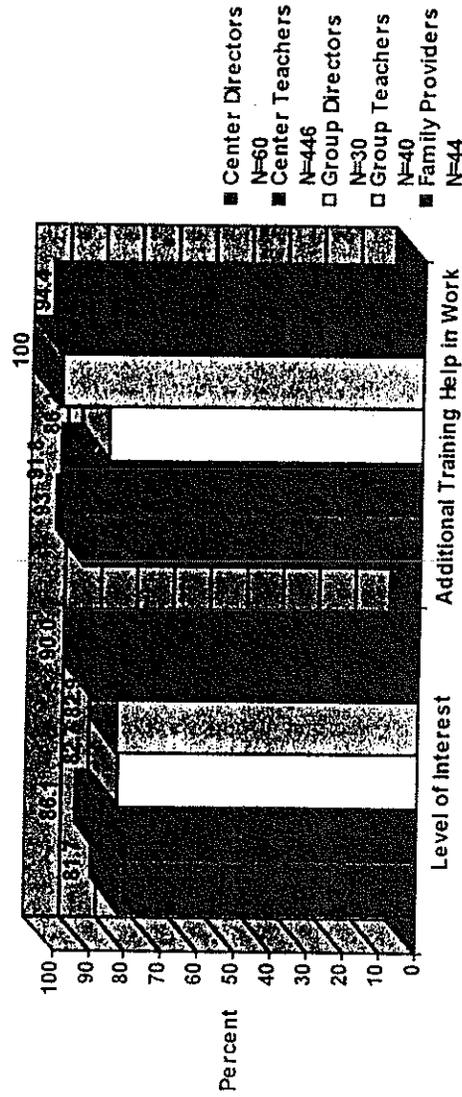
* Perceived helpfulness is indicated by the mean score for those who have experienced a method of training, on a scale of 1 = very helpful, 2 = somewhat helpful, and 3 = not helpful. In addition, a rank order of the methods is indicated in parentheses.

Lack of child care for their own children while attending training is considered important as a barrier only by family providers. This is another expected finding since family providers are most likely to have to attend training outside of work hours, necessitating the need to find care for their own children while attending training.

Contrary to what we might expect given the current lack of status and minimal reward system for child care providers, having **no long term gains or rewards** for training is not considered a very important barrier by provider groups. However, center directors, in comparison to other provider groups, were more likely to perceive this as an important barrier.

Having **no one to watch the children during the child care hours** is seen as the most significant barrier to training by all provider groups. Center directors and family child care providers, however, are more likely to indicate this as a very important factor than are teachers and group providers. This is to be expected, since directors and family providers are responsible for finding substitutes in their child care settings.

In identifying other barriers to training, we can also examine the reward system attached to training. Providers were asked, "Do you receive any compensation for attending relevant training?" Figure 4 shows the types of compensation received by the center and group child care providers. Few providers receive any type of compensation, i.e. being paid while in training, receiving compensatory time, or being reimbursed for expenses. Center directors appear to fare better than other provider groups – 55.9% indicate that they are paid while in training. This can be interpreted that they are more likely to attend relevant training during the work hours.



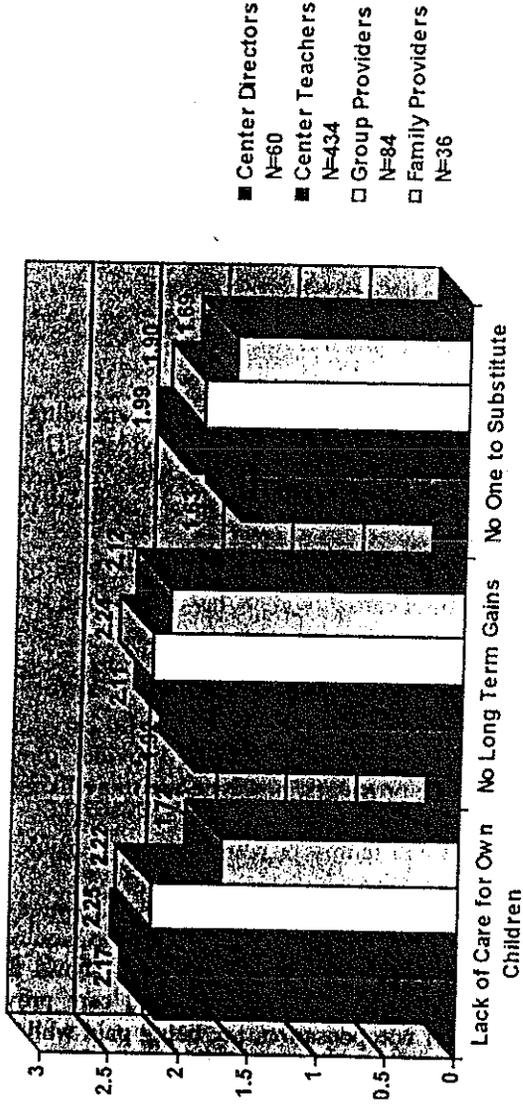
* For level of interest, numbers represent:

- The percentage of family providers, group teachers, and center teachers who indicate they are "interested" or "very interested" in taking workshops or training.
- The percentage of center directors and group directors who indicate their staff are "interested" or "very interested" in taking workshops or training.

For helpfulness of additional training in one's work numbers represent:

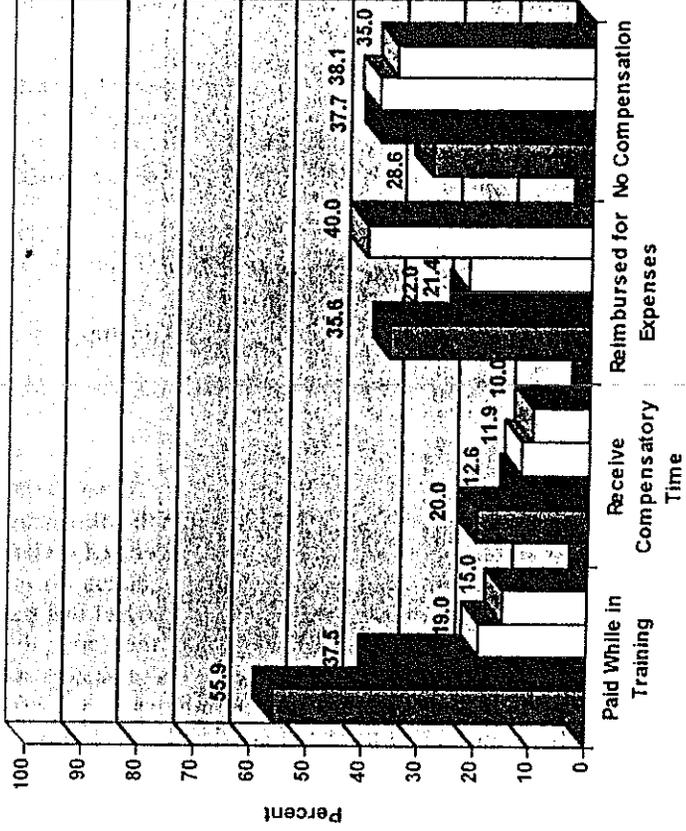
- The percentage of providers who indicate that attending additional workshops of training will "somewhat" or "very much" help them in their work.

Fig. 2. Level of Interest and Perceived Helpfulness of Additional Training.*



* Importance of barrier is indicated by the mean score of the provider group on a scale of 1-very important, 2=somewhat important, and 3=not important.

Fig. 3. Barriers to Training.*



* Percent indicating "yes" for each of these types of compensation that they receive for relevant training.

Fig. 4. Compensation for Attending Relevant Training.*

The Work Environment of Child Care Facilities

The quality of work life is not only an indicator of one type of quality within a child care environment, but it is an important factor that can influence the overall quality of care for young children, as well. As stated previously, we have used Jorde-Bloom's Early Childhood Work Environment Survey (ECWES) to assess a number of dimensions of the work environment within child care centers. The ECWES¹³ includes measurements of:

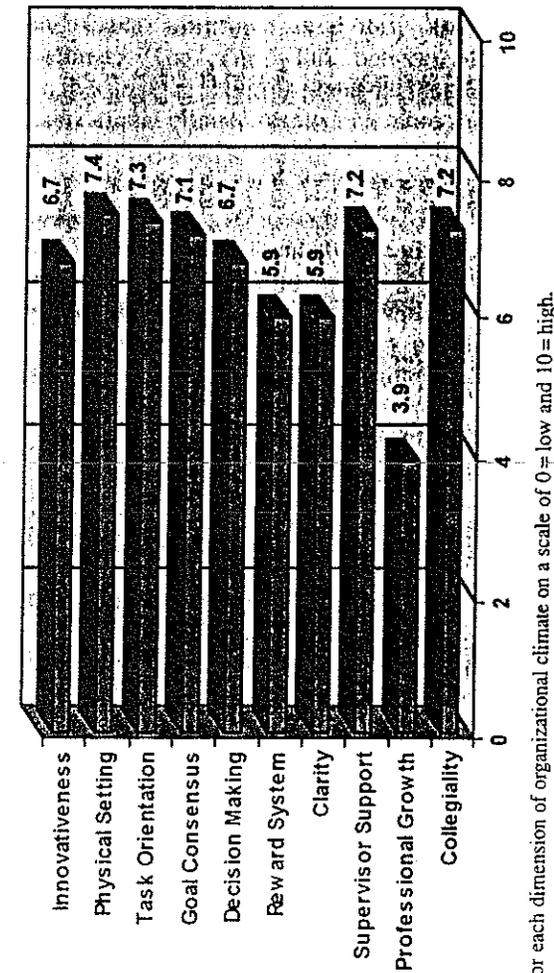
- Ten dimensions of *organizational climate* (collegiality, professional growth, supervisor support, clarity, reward system, decision making, goal consensus, task orientation, physical setting, and innovativeness);
- The importance that staff assign to each dimension (*summary of worker values*);
- The staff's *overall commitment* to the center;
- How the current *work environment resembles the staff's ideal*;
- The importance of various *educational goals and objectives*;
- The *degree of influence of the teaching staff* regarding various organizational dimensions.

Organizational Climate

The ten dimensions of organizational climate are shown in Fig. 5. In analyzing the scale values, which can range between 0 and 10, we see that the dimension of **professional growth** ranks at the bottom (3.94), followed by **reward system** (5.88) and **clarity** (5.91). This indicates that overall, staff in centers do not perceive many opportunities for professional growth, they do not feel that pay and fringe benefits are fair and equitably distributed, and they feel that communication about policies and procedures is unclear. These results are similar to national data where professional growth opportunities and reward systems are evaluated poorly by most child care staff (Jorde-Bloom, 1996).

It is important to determine what factors, if any, are associated with these ten dimensions of organizational climate. Table 8 provides a summary from an analysis of relationships between each of the organizational climate dimensions and a series of factors. A number of **director characteristics** are examined first. In addition, **characteristics of teachers** (aggregated per site) and overall **site characteristics** are analyzed.

Overall, the average age of teachers is significantly related to all dimensions of organizational climate. Centers with older workers have a more positive work environment. Correspondingly, two other factors that are closely related



* Mean value for each dimension of organizational climate on a scale of 0 = low and 10 = high.

Fig. 5. Organizational Climate at Child Care Centers.*

Table 8. Factors Associated with Organizational Climate.

Factor	Collegiality	Professional Growth	Supervisor Support	Clarity	Reward System	Decision Making	Goal Consensus	Task Orientation	Physical Setting	Innovativeness
Director Background										
Age	-0.005	0.27*	-0.09	0.02	0.10	-0.14	0.05	0.03	0.11	0.05
Educational level	0.01	0.17	0.15	0.08	0.14	0.03	0.02	0.17	0.14	0.14
Years in field	0.08	0.44***	0.20	0.18	0.30*	0.24	0.19	0.26*	0.30*	0.26*
Years in present job	0.26*	0.40**	0.23	0.17	0.40**	0.19	0.12	0.26*	0.30*	0.14
Full-time/part-time	-0.23	-0.08	-0.06	-0.18	-0.32*	0.03	-0.13	-0.11	-0.03	-0.16
Long term educational goal	-0.14	0.34**	-0.07	0.13	0.15	0.04	-0.12	0.03	0.03	-0.00
CDA Status	0.06	0.20	0.07	-0.02	0.01	-0.12	-0.03	0.00	-0.06	0.03
Training hours/year	-0.13	0.02	-0.05	0.15	-0.12	0.02	0.04	0.00	-0.03	0.13
Annual training goal	-0.01	0.18	-0.05	-0.08	0.08	-0.01	-0.02	0.12	-0.03	0.13
Aggregate Teacher Characteristics	-0.20	0.18	-0.14	-0.05	-0.06	-0.14	-0.07	-0.11	0.07	0.06
Average age of teachers/site	0.39**	0.45***	0.29*	0.28*	0.43***	0.26*	0.34**	0.38**	0.32**	0.29*
Average educational level	-0.03	0.25	0.04	0.16	-0.07	0.01	-0.03	0.09	-0.08	0.00
Average years in field	0.24	0.36**	0.18	0.30*	0.27*	0.12	0.17	0.18	0.04	0.10
Average years in present job	0.36**	0.38**	0.28*	0.26*	0.28*	0.25	0.26*	0.29*	0.17	0.20
% of teachers working full-time	0.04	0.10	-0.13	-0.04	0.00	-0.07	0.15	-0.11	-0.05	0.09
Average teacher salary	0.15	0.53***	0.13	0.31*	0.27*	0.15	0.23	0.29*	0.18	0.24
Average long-term educational goal	-0.10	-0.06	-0.07	-0.08	-0.13	-0.13	-0.02	-0.14	0.05	-0.09
% teachers with annual goal 12 + hrs	0.01	0.43***	-0.01	-0.02	-0.01	-0.06	0.06	0.05	0.08	0.14
Average training hours/year	-0.09	0.11	0.23	-0.22	0.09	-0.18	-0.13	-0.10	-0.01	-0.20
Average teacher interest in training	-0.00	0.00	-0.14	-0.06	-0.03	-0.07	-0.09	-0.11	-0.08	-0.06
Site Characteristics										
Licensed capacity	-0.08	0.18	0.07	-0.11	0.08	0.02	-0.11	0.01	0.22	0.11
Number of classrooms	-0.09	0.33**	-0.05	0.05	0.02	0.11	0.00	0.06	0.11	0.15
Turnover	-0.31*	-0.33*	-0.40**	-0.26	-0.20	-0.30*	-0.39**	-0.24	0.28*	-0.26
Accreditation status	-0.18	-0.33**	-0.16	-0.24	-0.36**	-0.13	-0.28*	-0.25	-0.29*	-0.20
Average training hours/year	-0.07	0.17	-0.16	-0.18	0.14	-0.13	-0.12	-0.06	0.03	-0.12

* p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001.

Professional Development and The Quality of Child Care

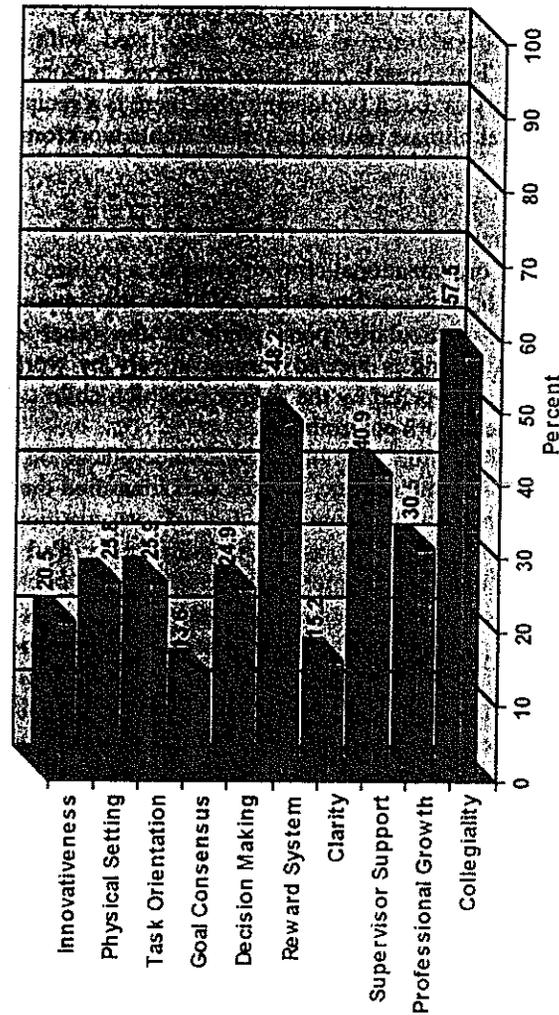
to each other (i.e. average number of years in current position and site turnover) are also significantly related to a number of the dimensions of organizational climate. In addition, a number of these organizational climate dimensions are more positive in centers that have older, more experienced directors. Hence, an older and more stable workforce is closely associated with a positive organizational climate. The causal link between these factors cannot be determined from this analysis, but it is plausible that there is a reciprocal effect – a positive organizational climate results in a more stable workforce and vice-versa.

Summary of Worker Values

The previous analysis of organizational climate gives us a picture of how child care centers fare on each of these dimensions. We see that opportunities for professional growth are particularly poor while at the other end of the continuum, the physical setting is viewed very positively by workers. These perceptions, however, are tempered by the degree to which child care workers value these aspects of their work environment.

Figure 6 gives us an indication as to the overall value placed on each of the 10 dimensions of organizational climate. Center staff identified the three most important aspects of their work from the list of organizational climate dimensions. The most highly valued aspect is **collegiality and co-worker relations** – over 60% of child care center staff identify this as one of the three most important aspects of their work. **Reward system** – fairness in pay and benefits – is second most important (48.2%) and **supervisor support** is third (40.9%) most important. The dimension of **opportunities for professional growth** comes in fourth with 30.5% of the caregivers identifying it as important. Therefore, even though staff do not perceive many opportunities for professional growth, this aspect is not as highly valued as other areas. Those areas least valued are **goal consensus** (13.5%), **clarity** (15.2%), and **innovativeness** (20.5%).

How these organizational climate dimensions are rated compared to the value placed on them gives us an indication where to focus improvement efforts. **Improvement in the reward system** will probably accrue the most lasting results since it is very poorly rated, yet highly valued. **Improving opportunities for professional growth** is also an area in which to focus attention since it is the most poorly rated area and ranked fourth in importance. On the other hand, **collegiality** is very important to workers, but given its positive assessment as a dimension of organizational climate, there is no need to improve it.



* Percent of center staff (averaged across sites) that identify each dimension of organizational climate as important.

Fig. 6. Summary of Worker Values.*

Summary of How Current Work Environment Resembles Ideal

As a way of understanding the discrepancy between ideal and real work conditions, child care workers were asked, "If you could design the ideal job, how close would your present position resemble this ideal position with respect to the following?" Responses range between *not at all like my ideal* to *is my ideal*. Based on this assessment, Figure 7 illustrates that the greatest discrepancy is in the **reward system**. There is a wide gap between what child care workers are paid versus what they think they should be paid. Given their current low salaries, this is an accurate appraisal on their part. The autonomy of staff to **make decisions** or express opinions on important issues is another area where child care staff feel that work environments least resemble the ideal.

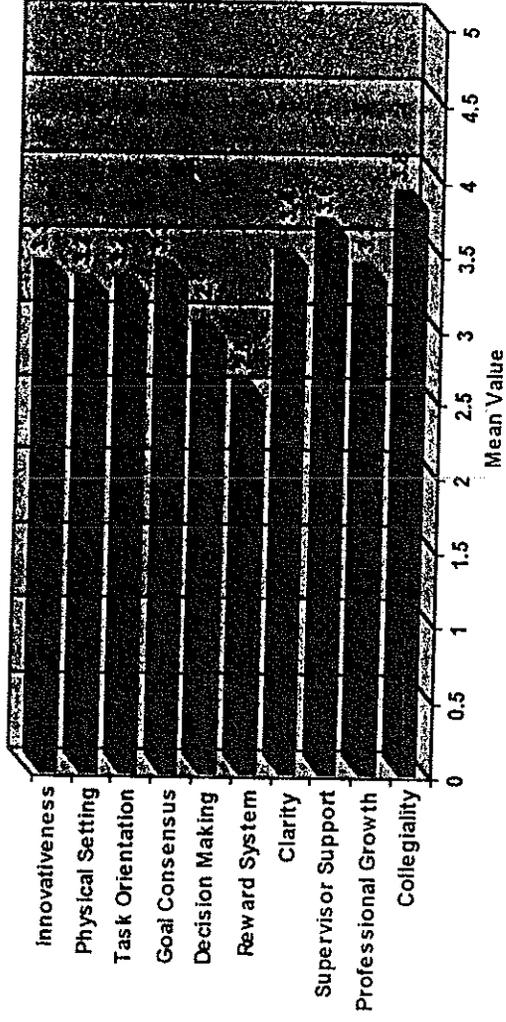
The smallest gap between the ideal environment and the real one experienced by child care workers is in the areas of **collegiality** and **supervisor support**. As far as **opportunities for professional growth**, the discrepancy between *ideal* and *real* falls mid-range on the continuum.

Importance of Educational Goals and Objectives

Early childhood programs can have a number of educational goals and objectives – but the priority given to each can vary across programs. Figure 8 shows how these educational goals and objectives are ranked in Pennsylvania child care centers. Consistent with developmentally appropriate practice in the early childhood field, the greatest emphasis is on **helping children to develop positive self concepts and self esteem** while the least emphasis is placed on helping children develop concepts needed for reading and math.

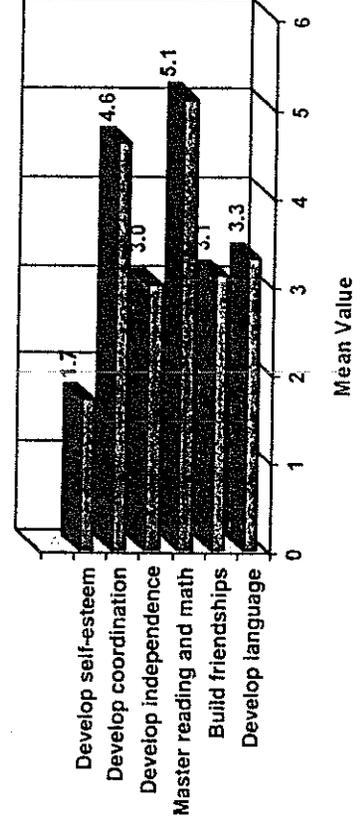
Degree of Influence of Teaching Staff Regarding Organizational Decisions

Perceptions of workers regarding the degree of influence of teaching staff with respect to various organizational decisions provides a fuller understanding of the decision making dimension of organizational climate. Staff were asked how much influence they have (very little to considerable influence) in ordering materials and supplies, interviewing and hiring staff, determining program objectives, training new aides or teachers, and planning daily activities. Figure 9 depicts the difference between what directors perceive is the degree of influence versus what teachers perceive is the degree of influence. Not unexpectedly, teachers do not perceive that they have as much influence as what directors say they do. This discrepancy also points to an area where improvement efforts can be focused.



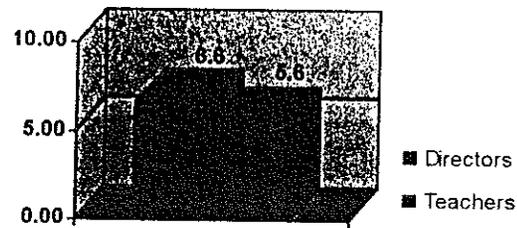
* Mean value for center staff on a scale of 1 = not like my ideal to 5 = like my ideal when asked "If you could design the ideal job, how close would your present position resemble your ideal work environment with respect to the following?"

Fig. 7. Summary of How Current Work Environment Resembles Ideal.*



* Mean value of importance on a scale of 1 = most important and 6 = least important.

Fig. 8. Importance of Organizational Goals.*



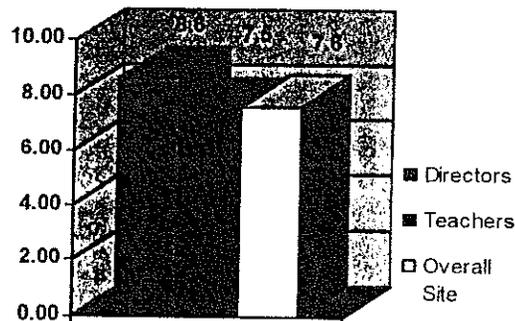
* Mean value on a scale of 0=very little influence to 10=considerable influence when asked, "How much influence does teaching staff currently have in each of the following areas below:

- ordering materials/supplies
- interviewing/hiring new staff
- determining program objectives
- training new aides/teachers
- planning daily schedule of activities."

Fig. 9. Degree of Influence of the Teaching Staff Regarding Various Organizational Decisions.*

Overall Commitment

All of the characteristics that have been discussed provide an understanding of specific areas where attention can be paid in intervention efforts to improve child care work environments. The commitment scale provides a summary of overall commitment of child care staff to their centers. Individuals who feel deeply committed tend to put extra effort into their work and take pride in their centers. In such environments, turnover is generally lower. Commitment among Pennsylvania child care staff is relatively high. Figure 10 indicates that



* Mean value on a scale of 1 = not committed to 10 = highly committed.

Fig. 10. Overall Commitment.*

directors have a slightly higher level of commitment than teachers – 8.6 for directors compared to 7.5 for teaching staff. Together this gives us an overall value of 7.6 for child care centers.

Quality of Care

Quality of Care as Assessed through the Environment Rating Scales

A previous section discussed results from observations of child care sites and the scale values for each individual item on the FDCRS, ITERS, and ECERS. That discussion focused on these data as a means to assess training needs. Our focus now shifts to the question about overall quality of care and its changes over the years, based on earlier studies conducted in Pennsylvania in 1984 (Kontos & Fiene, 1987) and in 1989 (Fiene & Melnick, 1991).

Overall, average ECERS scores have improved through the years, although the changes are not statistically significant, increasing from 3.78 in 1984 to 4.27 in 1989 to 4.63 in 1996. An analysis of FDCRS scores shows a marked (statistically significant, $p < 0.05$) improvement from 1989, increasing from 3.80 in 1989 to 4.47 in 1996. Several observations can be made in comparing 1996 with 1989 and 1984 data sets. While program quality scores on the ECERS and FDCRS have improved over the 12-year time frame, the bad news is that quality scores, on average, are still at the mediocre level. The ITERS score in 1996 is even worse (4.23) and is a major concern. There are no comparable data for the ITERS from the 1989 or 1984 research studies. Overall, Pennsylvania child care has improved, but it is still not in the good or excellent range. National and international data from research studies are very similar with ranges from 3.70 for family child care homes with little training to 5.22 for child care centers that are accredited (Phillips, 1987).

What are some reasons for the improvements? Two major interventions occurred during this 12-year time period. Both occurred at approximately the same time so it is difficult to determine the contribution of each to the overall improvement in quality. In 1992 new child care regulations were promulgated and the new comprehensive PA CC/ECD Training System was implemented. New regulations were an improvement over existing regulations, but the regulations deal primarily with basic health and safety issues. Although this will contribute to overall quality, it will not be a major contributor (Fiene & Melnick, 1991). What has been and continues to be a major contributor is the training system that has been implemented. When data are compared from the ECERS and FDCRS, family child care homes improved significantly more than child care centers. The home-based training system has been in place for twice as long as the center-based system. This is a very encouraging result.

Factors Associated with the Quality of Child Care

The previous analysis presents an overall picture of the quality of child care in Pennsylvania and the progress made in improving quality. Analysis presented in this section examines the current data to determine what factors are significantly related to the quality of care. Part of this analysis will be based on a data set that has matched the environment rating scales with the child care providers that were observed. This data set establishes the most direct link between an indicator of quality and the set of factors that might be associated with it (e.g. background characteristics of the caregiver, level of training of the caregiver, and caregiver's assessment of organizational climate). Other parts of this analysis will be based on a site level data set where aggregate values for most of the variables have been created to represent the site, overall. Where necessary, data have been weighted to adjust for the different probabilities of sample selection (i.e. the FDCRS included both family and group homes while the ECERS included both centers and group homes, necessitating that these analyses be based on weighted data). This analysis, which will include both bivariate correlation and multivariate regression analyses, will be presented separately for each of the three environment rating scales.

Table 9 provides a summary of results of a series of bivariate correlations between the measure of quality (i.e. either the FDCRS, ITERS, or ECERS average score) and a set of factors hypothesized to be related to quality (e.g. caregiver background characteristics, training experience, and assessment of organizational climate; for family providers, in lieu of organizational climate, an indicator of their *connectedness* to a child care network is used).

Bivariate analysis of the FDCRS finds four factors that are significantly correlated with quality of family child care. Family caregivers that are younger and have higher long-term educational goals are more likely to provide a higher quality of care. The other factors are measures of a family provider's assessment of the current training system. Providers who evaluate the current system of training as **inappropriate to their skill level** and **not useful for their work** as a family caregiver are more likely to provide a higher quality of care. This is not as unexpected as it sounds. It is likely that those providers who are already providing quality care do not find as much benefit from the current training system that focuses most of its attention on entry level skills. In an attempt to further analyze this unusual finding, we examined the relationship between hours of training and evaluation of training by providers. We see that providers who have more hours of training are also more likely to rate the current system positively in terms of **goal achievement** ($B = -0.24, p \leq 0.07$), **appropriate skill level** ($B = -0.28, p \leq 0.03$), and **usefulness** ($B = -0.45, p \leq 0.000$).¹⁴

Table 9. Factors Associated with Quality of Care.

Factor	FDCRS (N=67)†	ITERS (N=36)	ECERS (N=57)†
Caregiver Background			
Age	-0.30**	0.13	0.05
Educational level	0.003	0.14	0.11
Years in field	-0.20	0.25	0.04
Salary	0.19	0.43**	0.36**
Long term educational goal	0.27*	0.07	-0.09
Training Characteristics			
Annual educational goal	0.12	0.19	0.14
Training hours per year	-0.03	0.02	-0.09
CDA status	0.05	0.005	0.03
Training helpful in work	0.08	-0.16	-0.03
Evaluation of training system			
a. appropriateness	0.29**	-0.01	0.03
b. goal achievement	0.11	0.07	-0.10
c. usefulness	0.28**	0.08	-0.05
d. applicability	-0.02	-0.16	-0.08
Organizational Climate			
	N.A.††		
Overall commitment		0.01	0.14
Collegiality		-0.31	-0.005
Professional growth		0.15	0.41**
Supervisor support		-0.29	0.22
Clarity		0.06	0.32*
Reward system		-0.27	0.29*
Decision making		-0.24	0.17
Goal consensus		-0.18	0.32*
Task orientation		0.05	0.30*
Physical setting		0.20	0.21
Innovativeness		-0.07	0.23
Connectedness			
(Family Child Care Only)	-0.10	N.A.	N.A.

† This is the weighted N since there were observations made in more than one type of child care (i.e. family, group, or center).

†† N.A. = Not applicable

* $p \leq 0.05$; ** $p \leq 0.01$.

Our bivariate analysis of the ITERS finds only **salary level** of the caregiver to be significantly related to quality of infant/toddler care. Caregivers with higher salaries provide higher quality infant/toddler care. The bivariate analysis of the ECERS reveals a number of factors that are significantly correlated with

quality of child care: salary and organizational climate factors of **professional growth, clarity, reward system, goal consensus, and task orientation**. Thus, the caregivers that provide high quality early childhood care are more likely to:

- have higher salaries;
- indicate that their center has opportunities for professional growth;
- feel that communication at their center is good and that work schedules, job descriptions, and rules are clear and well-defined;
- indicate that the pay and fringe benefits are fair and equitably distributed in their center;
- indicate that staff at their center agree on school philosophy, are united in their approach, and are committed to program goals and objectives;
- believe that they work hard but still have time to relax, that program procedures are efficient, and that meetings are productive.

In an analysis of the site level data set, we created an overall *quality of care* variable as an indicator of child care quality.¹⁵ In a multivariate analysis of these data, we then determined what site level factors significantly contribute to the variance in *quality of care* at the site level. Initially, we did not include any of the work environment variables (Bloom items) since that would result in excluding all home-based providers from the analysis. Our analysis reveals that **size** and **turnover** are significant factors and explain 19% of the variance in quality of care. Results from the regression analysis are: **class number** ($B=0.11$, $p \leq 0.0053$) and **turnover** ($B=-1.08$, $p \leq 0.0216$). Thus, sites with more classrooms and lower turnover have a higher quality of care.

When we add the Bloom items on organizational climate (thereby eliminating all home-based providers from the analysis), we find that only **opportunity for professional growth** independently contributes to the variation of quality of care at child care sites (the majority of which are centers). The results from the regression analysis for **professional growth** are $B=0.45$, $p \leq 0.0000$. Forty percent of the variation is explained by this factor. As we hypothesize, child care facilities that have more opportunities for professional growth have a higher overall quality of care.

CONCLUSIONS AND POLICY IMPLICATIONS

What can we conclude as a result of this research and what are the implications for public policy? There are numerous issues addressed and volumes of data analyzed. First, we can examine the overall conclusions with regard to the current training system – how it is evaluated and what the training needs are.

Overall, the training system is viewed positively by the provider groups, as evidenced by their response to questions about appropriateness, usefulness, applicability, and effectiveness in achieving learning objectives. Furthermore, providers express a high level of interest in the training and, for the most part, feel that additional training will help them in their work. When directors of centers and home-based providers are asked about the particular methods of training that they perceive to be most helpful to them and their staff, training methods that provide direct contact with a trainer (e.g. on-site training and workshops) are viewed as most helpful. Methods where the contact is indirect (e.g. video and satellite training) are viewed as least helpful. This is understandable, although the policy implications of this are not to abandon some of the more indirect, yet very cost effective methods of training like the learn-at-home videos and satellite training. A balance of methods is important in a system as massive as this one, where the ability to access training varies tremendously across providers and where resources are limited.

Regarding training needs, there is a high degree of consistency across provider groups in terms of areas they perceive as most critical. They identify supervision/discipline of children, social development (dealing with conflict), child development, and developmentally appropriate practice as areas with the highest priority. Although, providers do not identify any topic area as *not a priority* for training.

On the other hand, if we use the environment rating scales as an indicator where there are weaknesses in child care settings (hence, an area in need of training), we see a slightly different picture. Given the overall low score for the infant/toddler area (ITERS), any training in this area can be viewed as a priority. In addition, these items are consistently ranked low on all three environment rating scales: cultural awareness, personal grooming, dramatic (pretend) play, and sand and water play; furthermore, these areas are rated low in two out of the three environment rating scales: displays for children (FDCRS and ITERS), space alone (FDCRS and ECERS), helping infants/toddlers understand language (FDCRS and ITERS), art (ITERS and ECERS), and blocks (FDCRS and ITERS). Indeed, a number of these items from the environment rating scales fall under the broader categories of social development, child development, and developmentally appropriate practice. The information from the environment rating scales offers more specific areas of need.

One particularly interesting finding is the contradictory information related to the area of supervision/discipline of children. While providers identify this as a high priority area for training, the environment rating scales indicate an assessment in the *good* range for the discipline item. Again, this shows that

providers are performing better in this area than they think and it reveals the extent to which this is viewed as one of the most challenging areas in child care.

Turning now to the issue of quality of care and the factors that are associated with it, we find that our results that examine the relationship between the level of training at a site and quality of care are not as predicted. We do not find that the number of hours of training is a significant predictor of quality. What we do see, however, is that the most significant change in the quality of care since 1989 has occurred in family child care sites. Although we are unable to definitively conclude that the training system has been instrumental in improving quality of care in family homes, we do note that the home-based training system has been in existence for the longest period of time – over 10 years. Furthermore, the intervention effort (i.e. hours of training per year that staff at a site average), is still considerably low – 8.5 hours on average, with 98% of sample sites averaging fewer than 18 hours per year. A threshold for training to show some impact is around 18 hours according to other research (Howes, Smith & Galinsky, 1995). Given this, it is not unexpected that we do not find a significant relationship between number of hours of training and quality of care – there simply is not enough intervention (i.e. training hours) to determine impact.¹⁶ What does this mean for public policy? While limitations of the research design and measurements do not allow for definitive conclusions, there are some tentative policy implications that can be drawn from this study (Kagan & Wechsler, 1998). One concerns the number of hours of training that are mandated in the state regulations for child care. It indicates the need to increase the number of hours of training for child care providers if a significant impact of training is to be detected.

This is further supported when we see the strong association between the organizational climate dimension of opportunities for professional growth and overall quality of care at the site level. Centers where staff report more opportunities for professional growth have a higher quality of care and this factor, alone, explains a considerable portion of the variation in quality (40%). This finding substantiates the importance of fostering professional growth opportunities for child care providers. But it also implies the importance of making sure that these opportunities are linked to a model of career development and progression – not just a few hours of training that providers haphazardly take because they have to or because they are offered at a time that fits their schedule.

Overall, these data have given us some solid evidence to guide the development of the training system in Pennsylvania. We have highlighted some very specific areas where there is a need for training and we have shown the

clear association between opportunities for professional growth and the quality of care. Although there are some anomalies in the data and some unexpected findings, as a whole, these data are supportive of the efforts to implement a training system that fosters career development in the prediction that these efforts will improve the quality of care for children in Pennsylvania.

NOTES

1. This included a set aside of 6.25% of \$731,915,000 in 1991 federal funds for program quality initiatives.
2. A decision was made not to extend the data collection process for one additional family site after we had difficulty in scheduling the final site visit.
3. There were 26 paired observations analyzed to determine inter-rater reliability. For the ECERS, the rank order correlation was 0.90; for the ITERS it was 0.95; for FDCRS, it could not be calculated since there was only one paired observation; however, a visual inspection of the FDCRS data shows a high degree of consistency across observers.
4. The instruments include specific descriptions of what to look for in assigning a value of 1,3,5 and 7 for each of the items assessed. A mid-point rating of 2,4 or 6 is given when all the lower and part of the higher description applies. The internal consistency scores (Cronbach's Alpha) for each of the environment rating scales is 0.83 for ECERS, 0.83 for ITERS, and between 0.70 and 0.93 for the individual subscales of the FDCRS.
5. The total scale alpha coefficient for internal consistency for the ten dimensions of organizational climate is 0.95. The specific details on the reliability and validity for other components of ECWES can be found in P. Jorde-Bloom (1996), *Improving the Quality of Work Life in the Early Childhood Setting: Resource Guide and Technical Manual for the Early Childhood Work Environment Survey*. Wheeling, Illinois: The Early Childhood Professional Development Project.
6. They were asked to indicate if there is a need for training, based on a scale of 1 to 4 where 1 = a very serious need and 4 = not a training priority.
7. For the purpose of this analysis on training needs, both the directors and teachers within a group home have been combined into one category, representing group child care providers. This decision was made because the child care setting is usually small and a distinction cannot always be made between a "director" and "teacher" within the group site.
8. For example, the ECERS doesn't assess health and safety areas and the discipline item is spread across a number of supervision items.
9. Each factor was assessed by providers as 1 = very important, 2 = somewhat important, or 3 = not important.
10. It is important to keep in mind that when providers indicate that they learned the material, this is based on their subjective assessment, and the extent to which they actually did learn the material is not objectively measured through this question.
11. Teachers in both group and center settings were asked, "In general, how interested are you in taking workshops or courses on teaching and/or caring for

children?" Directors were asked to indicate their perception of interest on the part of their staff.

12. Family providers were asked, "Based on your experience, what method(s) of training are most helpful for you?" Center directors, were asked, "Based on your experience, what method(s) of training are most helpful for your staff?"

13. The analysis in this section only includes child care centers since we used Bloom's instrument only in facilities that had more than three staff, as recommended. There were a total of 60 centers included in the data set, however, due to missing data from some centers, only 55 are included in the analysis presented herein.

14. These coefficients are negative since a lower value on each of the evaluative factors indicates a more positive assessment.

15. If there were two classrooms observed, the new overall *quality* measure was an average of the two scores (regardless of the type of classroom observed). If only one classroom was observed, then that score became of the site's overall *quality* score.

16. There may also be measurement problems related to the *quantity of training* variable. Issues related to recall on the part of the provider and definitions of what constitutes training may vary – both of which can affect the reliability and validity of the reported hours of training.

ACKNOWLEDGEMENTS

Keystone University Research Corporation would like to thank the many individuals who contributed to the completion of this study. First, the Department of Public Welfare, Office of Children, Youth, and Families, Bureau of Child Day Care is commended for their commitment to child care and efforts to improve quality. Without their funding of this project, it could not have been completed. The support of Richard Fiene, Ph.D., the DPW Project Officer, has been particularly instrumental in moving this research forward and guaranteeing that the policy implications of the findings are not lost. In addition, others within DPW – Kathryn Holod, Director, Bureau of Child Day Care and Jo Ann R. Lawer, Esq., Deputy Secretary for Children, Youth, and Families – are thanked for their continued support of this effort.

Thanks are also extended to the family, group, and center child care facilities that participated in this study as sample sites. Without their willingness to participate in this study and to be open and honest with the fieldworkers, this study would not have been possible. Furthermore, Lisa Heintzelman, Carla Thompson-John, Patricia Robinson-Burns, Ronnie Braun, Debora Reiff, Robin Eckert, Judy Peterson, Barbara Marchese, Majida Mehana, Kristi Hannan, Winifred Feise, Marilyn Albert, Jeanette Twombly, and Genevieve Mann are thanked for their quality efforts as fieldworkers for this study.

Finally, the individuals who contributed countless hours to assist the principal investigators in implementing this project – Pauline Kraus and Scott Johnson of Keystone University Research Corporation – are thanked for their

perseverance over the long haul. The tasks of organizing the data collection process, monitoring the work of fieldworkers, organizing the data sets, and assisting in its analysis were complicated and required considerable attention to detail.

The overarching goal of this study is to contribute to our knowledge about how to improve the quality of care for children in Pennsylvania. We hope that the knowledge gained as a result of this study will take us one step further in that direction.

REFERENCES

- Barbour, N. E., Peters, D. L., & Baptiste, N. (1995). The child care associate credential program. In: S. Reifel (Ed.), *Advances in Early Education and Day Care (Vol. 7): Social Contexts of Early Development and Education*. Greenwich, CT: JAI Press.
- Bredenkamp, S., & Willer, B. (1994). Introduction. In: J. Johnson & J. McCracken (Eds.), *The Early Childhood Career Lattice: Perspectives on Professional Development*. Washington, D.C.: National Association for the Education of Young Children.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Caffarella, R. (1994). *Planning programs for adult learners: A practical guide for educators, trainers, and staff developers*. San Francisco, CA: Jossey-Bass.
- Children's Defense Fund (1998). *The state of America's children*. Washington, D.C.: Author.
- Children's Defense Fund (1994). *The state of America's children*. Washington, D.C.: Author.
- Chung, A., & Stoney, L. (1997). *Using research to improve child care for low-income families*. Washington, D.C.: Child Care Bureau, Administration for Children, Youth and Families, Department of Health and Human Services.
- Fiene, R. (1995). Utilizing a statewide training system to improve child day care quality. *Child Welfare*, 74(6), 1189-1201.
- Fiene, R., & Melnick, S. (1991). Quality assessment of early childhood program: A multi dimensional approach. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.
- Galinsky, E., Howes, C., Kontos, S., & Shinn, M. (1994). *The study of children in family child care and relative care: Highlights of findings*. New York, NY: Families and Work Institute.
- Helburn, S. (Ed.) (1995). *Cost, quality, and child outcomes in child care centers*. Denver, CO: Center for Research in Economics and Social Policy, Department of Economics, University of Colorado.
- Holt-Reynolds, D. (1995). Preservice teachers and coursework: When is getting it right wrong? In: M. O'Hair & S. Odell (Eds.), *Educating Teachers for Leadership and Change*. California: Corum Press.
- Howes, C., Smith, E., & Galinsky, E. (1995). The Florida child care quality improvement study: Interim report. New York, NY: Families and Work Institute.
- Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.
- Jones, E. (Ed.) (1993). *Growing teachers: Partnerships in staff development*. Washington, D.C.: National Association for the Education of Young Children.

- Jorde-Bloom, P. (1988). Assess the climate of your center: Use the early childhood work environment survey. *Day Care and Early Education*, (Summer), 9-11.
- Jorde-Bloom, P. (1998). Using climate assessment to improve the quality of work life in early childhood programs. In: S. Reifel (Ed.), *Advances in Early Education and Day Care* (Vol. 10). Greenwich, CT: JAI Press.
- Jorde-Bloom, P., & Sheerer, M. (1992). The effect of leadership training on child care program quality. *Early Childhood Research Quarterly*, 7, 579-594.
- Kagan, S. L., & Wechsler, S. (1998). Changing realities - changing research. In: S. Reifel (Ed.), *Advances in Early Education and Day Care* (Vol. 10). Greenwich, CT: JAI Press.
- Katz, L. (1994). Perspectives on the quality of early childhood programs. *Phi Delta Kappan*.
- Kontos, S., & Fiene, R. (1987). Child care quality, compliance with regulations, and children's development: The Pennsylvania study. In: D. Phillips (Ed.), *Quality in Child Care: What Does Research Tell Us?* Washington, D.C.: National Association for the Education of Young Children.
- Lewin, K. (1935). *A dynamic theory of personality: Selected paper*. New York: McGraw-Hill.
- Love, J., Schochet, P., & Meckstroth, A. (1996). *Are they in any real danger? What research does - and doesn't - tell us about child care quality and children's well being*. Princeton, NJ: Mathematica Policy Research, Inc.
- McAllister, J. (1997). The Children's Crusade. *Time*, (August 25), 36.
- Modigliani, K. (1993). Readings in family child care professional development: Project-to-project compiled. Boston, MA: Wheelock College Family Child Care Project.
- Morgan, G., Azer, S., Costley, J., Elliott, K., Genser, A., Goodman, I., & McGimsey, B. (1994). Future pursuits: Building early care and education careers. In: J. Johnson & J. McCracken (Eds), *The Early Childhood Career Lattice: Perspectives on Professional Development*. Washington, D.C.: National Association for the Education of Young Children.
- Morgan, G., Costley, J. B., Genser, A., Goodman, I. F., Lombardi, J., & McGimsey, B. (1993). *Making a career of it: The states of the states report on career development in early care and education*. Boston: The Center for Career Development in Early Care and Education, Wheelock College.
- Phillips, D. A. (1987). *Quality in child care: What does research tell us?* Washington, D.C.: National Association for the Education of Young Children.
- Phillips, D. A., Howes, C., & Whitebook, M. (1992). The social policy context of child care: Effect on quality. *American Journal of Community Psychology*, 20, 25-51.
- Stoney, L., Elliott, K., Chung, A., Genser, A., & Raggozzine, D. (1997). *Common Threads: Weaving a Training and Career Development System for 21st Century Pennsylvania*. Boston, Massachusetts: The Center for Career Development in Early Care and Education, Wheelock College.
- U.S. Department of Education (1990). *A profile of child care settings: Early Education and care in 1990*. A final report by Mathematica Policy Research, Washington, D.C.: author.
- VanderVen, K. (1994). Professional development: A contextual model. In: J. Johnson & J. McCracken (Eds), *The Early Childhood Career Lattice: Perspectives on Professional Development*. Washington, D.C.: National Association for the Education of Young Children.
- Willer, B. (1994). A conceptual framework for early childhood professional development. In: J. Johnson & J. McCracken (Eds), *The Early Childhood Career Lattice: Perspectives on Professional Development*. Washington, D.C.: National Association for the Education of Young Children.

Early experiences of adversity such as abuse are associated with serious short- and long-term negative consequences (Anda et al., 2006; Fantuzzo et al., 2011; Manly et al., 2001; Pears et al., 2008).

Though 0–5 year olds account for 46% of the >680,000 children identified annually as victims of maltreatment (U.S. DHHS, 2017), early care and education (ECE) providers report just 0.6% of these.

To the extent this inactivity indicates underreporting, it emphasizes ECE providers' crucial role for protecting young children from abuse.

Among the barriers to reporting suspected abuse (Alvarez et al., 2004; Dinehart & Kenny, 2015), one key obstacle is the lack of evidence-based training. A recent IOM report (2014) specifically identified the need to examine the efficacy of mandated reporter education programs.

Two small studies looked at in-person training for ECE providers (McGrath et al., 1987; Khan et al., 2005), and one evaluated an online curriculum (Kenny, 2007); but none evaluated scalable interventions.

The present study examined the efficacy of a scaled online educational program previously shown in a randomized controlled trial (n=741) to improve knowledge and attitudes about reporting suspected abuse (Mathews et al., 2017).

Methods

The present study employed a single group pretest-posttest design to measure changes among 5379 participants from Pennsylvania (see Table 1) regarding *knowledge* about ECE providers' responsibilities to report suspected abuse (23 items, correct/incorrect), and *attitudes* toward reporting suspected abuse (13 items, 7-point Likert scale).

Created specifically for ECE providers, the *iLookOut* online learning program employs an interactive, video-based storyline where learners take the role of an ECE provider faced with decision points on how to respond to indicators of potential abuse

Table 1

Demographics for the study sample (n=5379)

	%
Ethnicity	
Black/African-American	19.5
White	71.4
Hispanic	5.2
Asian	1.6
Native Hawaiian/Pacific Islander	.1
American Indian/Alaska Native	.4
Other	1.8
Parent/guardian	
Yes	60.2
Prior mandated reporter training	
Yes	74.4
Work setting	
Rural	17.9
Suburban	45.2
Urban	36.9
Education level	
Below High School	.7
High School or GED	31.9
Child Development Associate (CDA)	10
Associate's Degree	14.8
Bachelor's Degree	31.6
Masters or Doctoral	11
Gender	
Female	90.5
Age	
18 - 29	39.4
30 - 44	28.8
>44	31.8
Primary job responsibilities	
Teacher/caregiving staff	64.2
Early intervention specialist	1.4
Support staff	4.8
Director/Assistant Director	10
Other	19.7

Resource files provide information about child abuse, and guidance about when to report. For this Phase 2 open enrollment trial, ECE providers could access the *iLookOut* learning program through publicly available websites, but participants were not actively recruited. Participants provided informed consent prior to the registration/pre-test, and earned 3 hours of professional development credit by completing the learning program, post-test, and evaluation. No other incentives or remuneration were provided.

Results

One-way repeated measures ANOVA were conducted to compare the effects of *iLookOut* participation on pretest and posttest measures of knowledge and attitude.

Testing for main effects only yielded significant results for knowledge: Wilks' $\lambda = 0.508$, $F(1,5433) = 5256.964$, $p < 0.001$; partial $\eta^2 = 0.492$ and attitude: Wilks' $\lambda = 0.576$, $F(1,5432) = 4000.369$, $p < 0.001$; partial $\eta^2 = 0.424$ (see Fig. 1).

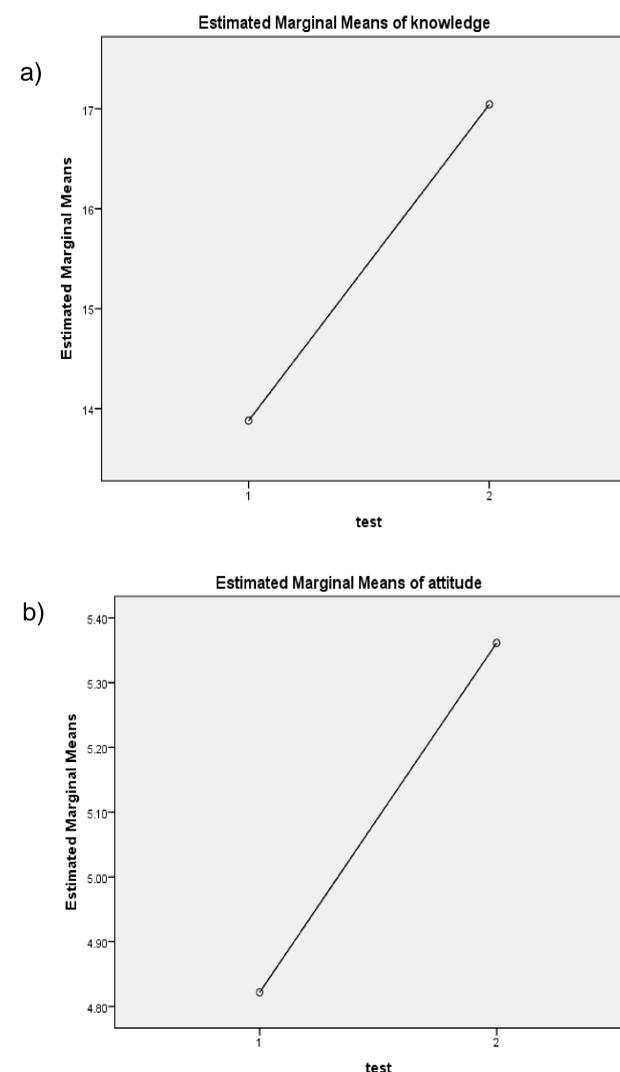


Figure 1. Mean difference between pre- and post-test on knowledge (a) and attitude (b) items. Significant at $p < 0.001$.

There was a significant interaction effect for knowledge and prior training on mandated reporting: Wilks' $\lambda = 0.995$, $F(1,5375) = 27.603$, $p < 0.001$; partial $\eta^2 = 0.005$ (see Fig. 2). Main effects of the online training on knowledge performance were still significant: Wilks' $\lambda = 0.554$, $F(1,5375) = 4335.154$, $p < 0.001$; partial $\eta^2 = 0.446$. This highlights gains on posttest performance even though participants had prior training in mandated reporter

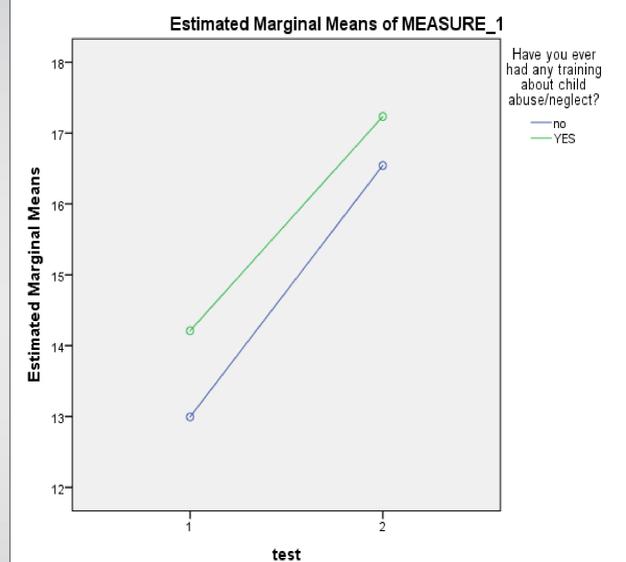


Figure 2. Mean difference between pre- and post-test on knowledge for participants with and without prior training

Discussion

Results from this study yielded the similar efficacy and effect sizes as did the initial randomized controlled trial of *iLookOut*.

They also demonstrate the feasibility of scaling an evidence-based, online mandated reporter training. Providing ECE providers open access to such training could help identify and prevent maltreatment of young children.

A 5-year randomized controlled trial is currently underway (1-R01-HD088448-02) to evaluate the impact of *iLookOut* on actual reporting practices of ECE providers.