

## THEORY OF REGULATORY COMPLIANCE

Richard Fiene

October 2016

*The Theory of Regulatory Compliance (TRC)*<sup>1</sup> deals with the importance and significance of complying with rules or regulations. This theory has implications for all rule, regulatory, and standards development throughout human service and economic domains although the research is being drawn from the human services field. The TRC has developed over the past 40 years. It has particular significance now as the need for either more or less oversight has become politically charged. What is important about the TRC is its emphasis on selecting the right rules rather than having more or less rules and the nature of these rules as being significantly predictive of positive outcomes by being in compliance with said rules.

The Theory of Regulatory Compliance was first proposed in the 1970's when the relationship between compliance with rules was compared to compliance with best practice standards and outcome data. From this comparison, it became clear that as facilities were in 100% compliance with all rules, their overall best practice scores and positive outcomes began to drop off. It was also found that there was a "sweet spot" at a substantial compliance level where best practice scores and positive outcomes were at their highest levels. In statistical terms, the relationship was *curvilinear rather than linear*. This initial result has been confirmed many times over the past 40 years in different forms of human service facilities. This result also led to the conclusion that possibly being in "full" or 100% compliance with all rules was not necessarily a good policy and that *all rules or regulations are not created equal*.

This led to the development of two methodologies dealing with risk assessment and key indicators of regulatory compliance. In both of these methodologies, the focus is on identifying a more targeted group of rules that either statistically predict overall regulatory compliance or reduce risk.

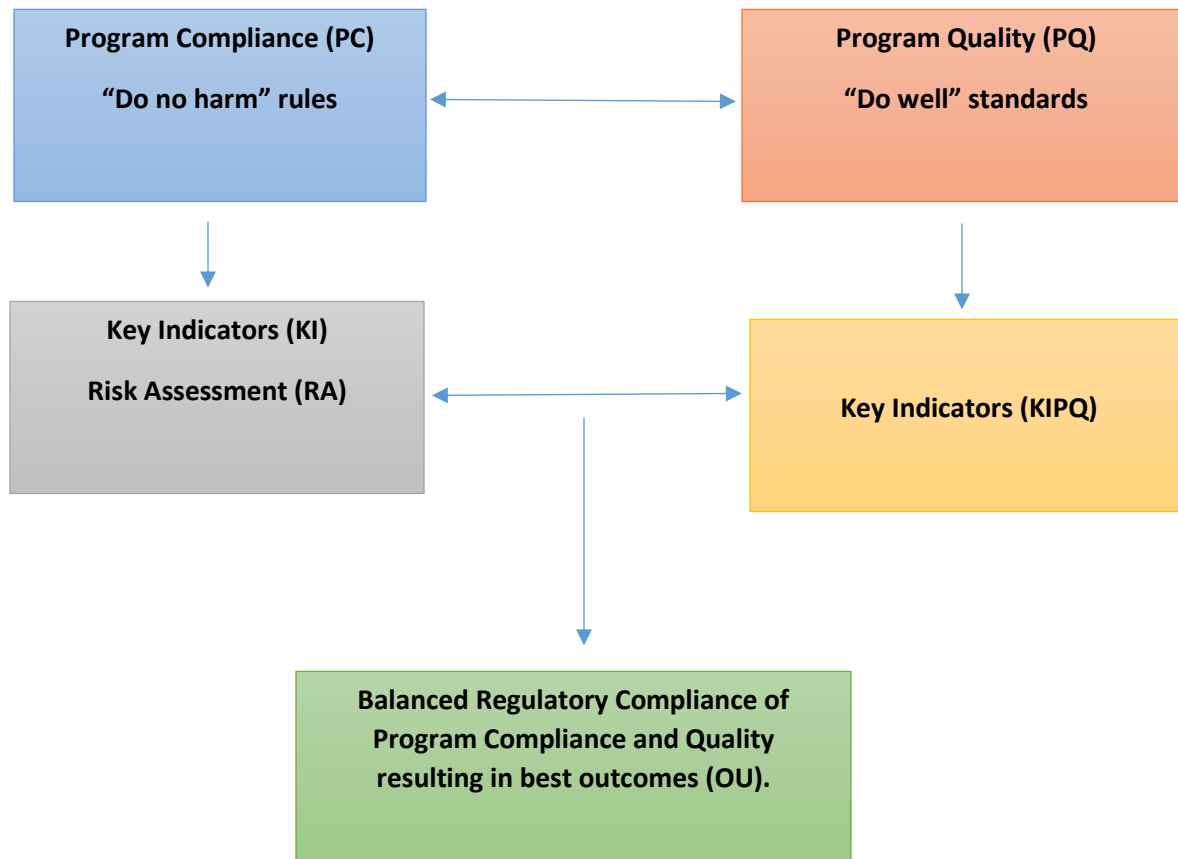
But what is the underlying reason for the TRC. It appears from data collected in various regulatory systems that the nature of the rules themselves may be the real problem. When rules are too minimal to comply with, it is *far more difficult to discriminate between the really good facilities and the mediocre facilities*. This unfortunately is the nature of regulatory data, it is dramatically *skewed data* with the majority of facilities being in compliance with all the rules.

The solution to the above dilemma is not to de-regulate or to over-regulate but to come up with the "right" balance of rules or regulations. We do not want to make the mistake of the old proverbial "throwing out the baby with the bathwater". We need to have some form of oversight but it needs to be the right balance of oversight based upon risk and predictive targeting of specific rules or regulations. The statistical methodologies exist to identify these specific risk and predictive rules and regulations.

- 
1. For additional information regarding TRC, please go to the following website: <http://RIKInstitute.com/RIKI>.
- 

**Richard Fiene, Ph.D., Research Psychologist, Research Institute for Key Indicators; Senior Consultant for Licensing Measurement, National Association for Regulatory Administration; and Affiliate Professor, Penn State Prevention Research Center.**  
[RFiene@NARALicensing.org](mailto:RFiene@NARALicensing.org)

### Theory of Regulatory Compliance Algorithm (Fiene, 11/16)



Balance of “do no harm” rules with “best practice” standards selected by risk and ability to predict positive outcomes. The Theory of Regulatory Compliance deals with selecting the “right” rules and standards that have predictive validity and do no harm. It acknowledges that all rules and standards are not created equal and have a differential impact in a monitoring or licensing system. By following a differential monitoring approach of key indicators and risk assessment, the most cost efficient and effective system can be implemented. The Theory of Regulatory Compliance proposes policy based upon substantial but not full compliance (100%) with all rules. The following algorithm summarizes TRC:

$$(PC < 100) + (PQ = 100) \rightarrow KI (10-20\% PC) + RA (10-20\% PC) + KIPQ (5-10\% of PQ) \rightarrow OU$$

## **Theory of Regulatory Compliance Math Modeling (Fiene, 11/16)**

This presentation will provide key definitions, a legend and math modeling concepts related to the Theory of Regulatory Compliance. It builds upon the previous two presentations on an overview and algorithm for the Theory of Regulatory Compliance (TRC).

### Legend/Definitions:

R = Rules/Regulations

C = Compliance with rules/regulations

NC = Non-Compliance with rules/regulations

KI = Key Indicators of substantial (99%) compliance with all rules/regulations

CI = Comprehensive Instrument measuring compliance with all rules/regulations

RA = Risk Assessment measuring the relative risk of non-compliance with specific rules/regulations

DM = Differential Monitoring using Key Indicators and/or Risk Assessment

### Math Modeling:

$$\Sigma R = C$$

Summation of all rules equals compliance score.

$$KI > 0 = CI$$

If KI greater than zero, use comprehensive instrument for measuring compliance with all rules/regulations.

$$RA (NC\%) = CI$$

If RA has a pre-determined % on non-compliance, use comprehensive instrument for measuring compliance with all rules/regulations.

$$KI + RA = DM$$

Key indicators plus Risk Assessment equals a Differential Monitoring Approach.

$$TRC = 99\% + \phi = 100\%$$

Theory of Regulatory Compliance equals substantial compliance but not full compliance.

$$NC + C = CI$$

Non-Compliance plus Compliance with all rules/regulations equals the score on the comprehensive instrument.

$$(CI < 100) + (CIPQ = 100) \rightarrow KI (10-20\% CI) + RA (10-20\% CI) + KIQP (5-10\% \text{ of } CIPQ) \rightarrow OU$$

Where  $CI < 100$  is substantial compliance with all rules or the 99% rule,  $CIPQ = 100$  maximizing doing well,  $KI (10-20\% CI)$  is key indicators are generally 10-20% of all rules as well as risk assessment ( $RA (10-20\% CI)$ ),  $KIQP (5-10\% \text{ of } CIPQ)$  is the percent of standards taken from program quality that become key indicators of quality, and finally  $OU$  are positive outcomes or results.

# Theory of Regulatory Compliance Monitoring Paradigms

Richard Fiene

December 2016

This paper provides some key elements to the two dominating paradigms (Relative versus Absolute) for regulatory compliance monitoring based upon the Theory of Regulatory Compliance. See the table below for the key elements summarized for the Monitoring Paradigms followed by a more detailed description of each key element. These key elements are all inter-related and at times are not mutually exclusive.

## Regulatory Compliance Monitoring Paradigms

**Relative** <-----> **Absolute**

*Substantial* <-----> *Monolithic*  
*Differential Monitoring* <-----> *One size fits all monitoring*  
*Not all standards are created equal* <-----> *All standards are created equal*  
*Do things well* <-----> *Do no harm*  
*Strength based* <-----> *Deficit based*  
*Formative* <-----> *Summative*  
*Program Quality* <-----> *Program Compliance*  
*100-0 scoring* <-----> *100 or 0 scoring*  
*QRIS* <-----> *Licensing*  
*Non Linear* <-----> *Linear*

**Relative versus Absolute Regulatory Compliance Paradigm:** this is an important key element in how standards/rules/regulations are viewed when it comes to compliance. For example, in an absolute approach to regulatory compliance either a standard/rule/regulation is in full compliance or not in full compliance. There is no middle ground. It is black or white, no shades of gray. It is 100% or zero. In defining and viewing these two paradigms, this dichotomy is the organizational key element for this paper.

**Substantial versus Monolithic:** in monolithic regulatory compliance monitoring systems, it is one size fits all, everyone gets the same type of review (this is addressed in the next key element below) and is more typical of an absolute paradigm orientation. In a substantial regulatory compliance monitoring system, programs are monitored on the basis of their past compliance history and this is more typical of a relative paradigm orientation. Those with high compliance have fewer and more abbreviated visits/reviews while those with low compliance have more comprehensive visits/reviews.

*Differential Monitoring versus One Size Fits All Monitoring:* in differential monitoring (Relative Paradigm), more targeted or focused visits are utilized spending more time and resources with those problem programs and less time and resources with those programs that are exceptional. In the One Size Fits All Monitoring (Absolute Paradigm), all programs get the same type/level of review/visit regardless of past performance.

*Not all standards are created equal versus All standards are created equal:* when looking at standards/rules/regulations it is clear that certain ones have more of an impact on outcomes than others. For example, not having a form signed versus having proper supervision of clients demonstrates this difference. It could be argued that supervision is much more important to the health and safety of clients than if a form isn't signed by a loved one. In a relative paradigm, all standards are not created nor administered equally; while in an absolute paradigm of regulatory compliance, the standards are considered created equally and administered equally.

*“Do things well” versus “Do no harm”:* “doing things well” (Relative Paradigm) focuses on quality of services rather than “doing no harm” (Absolute Paradigm) which focuses on health and safety. Both are important in any regulatory compliance monitoring system but a balance between the two needs to be found. Erring on one side of the equation or the other is not in the best interest of client outcomes. “Doing no harm” focus is on the “least common denominator” – the design and implementation of a monitoring system from the perspective of focusing on only 5% of the non-optimal programs (“doing no harm”) rather than the 95% of the programs that are “doing things well”.

*Strength based versus Deficit based:* in a strength based monitoring system, one looks at the glass as “half full” rather than as “half empty” (deficit based monitoring system). Emphasis is on what the programs are doing correctly rather than their non-compliance with standards. A strength based system is non-punitive and is not interested in catching programs not doing well. It is about exemplars, about excellent models where everyone is brought up to a new higher level of quality care.

*Formative versus Summative:* relative regulatory compliance monitoring systems are formative in nature where there is an emphasis on constant quality improvement and getting better. In absolute regulatory compliance monitoring systems, the emphasis is on being the gate-keeper and making sure that decisions can be made to either grant or deny a license to operate. It is about keeping non-optimal programs from operating.

*Program Quality versus Program Compliance:* relative regulatory compliance monitoring systems focus is on program quality and quality improvement while in absolute regulatory compliance monitoring systems the focus is on program compliance with rules/regulations with the emphasis on full, 100% compliance.

*100 – 0 scoring versus 100 or 0 scoring:* in a relative regulatory compliance monitoring system, a 100 through zero (0) scoring can be used where there are gradients in the scoring, such as partial compliance scores. In an absolute regulatory compliance monitoring system, a 100% or zero (0) scoring is used demonstrating that either the standard/rule/regulation is fully complied with or not complied with at all.

*QRIS versus Licensing:* examples of a relative regulatory compliance monitoring system would be QRIS – Quality Rating and Improvement Systems. Absolute regulatory compliance systems would be state licensing systems. Many programs talk about the punitive aspects of the

present human services licensing and monitoring system and its lack of focus on the program quality aspects in local programs. One should not be surprised by this because in any regulatory compliance system the focus is on "doing no harm" rather than "doing things well". It has been and continues to be the focus of licensing and regulations in the USA. The reason QRIS - Quality Rating and Improvement Systems developed in early care and education was to focus more on "doing things well" rather than "doing no harm".

*Non-Linear versus Linear:* the assumption in both relative and absolute regulatory compliance monitoring systems is that the data are linear in nature which means that as compliance with standards/rules/regulations increases, positive outcomes for clients increases as well. The problem is the empirical data does not support this conclusion. It appears from the data that the relationship is more non-linear where there is a plateau effect with regulatory compliance in which client outcomes increase until substantial compliance is reached but doesn't continue to increase beyond this level. There appears to be a "sweet spot" or balancing of key standards/rules/regulations that predict client outcomes more effectively than 100% or full compliance with all standards/rules/regulations – this is the essence of the Theory of Regulatory Compliance – substantial compliance with all standards or full compliance with a select group of standards that predict overall substantial compliance and/or positive client outcomes.

As the regulatory administration field continues to think about the appropriate monitoring systems to be designed and implemented, the above structure should help in thinking through what these systems' key elements should be. Both paradigms are important, in particular contexts, but a proper balance between the two is probably the best approach in designing regulatory compliance monitoring systems.

---

*For additional information:*

Richard Fiene, Ph.D., Research Psychologist  
Research Institute for Key Indicators (RIKILLC) (<http://RIKInstitute.com>)  
[RIKI.Institute@gmail.com](mailto:RIKI.Institute@gmail.com) or [RFiene@NARALicensing.org](mailto:RFiene@NARALicensing.org)  
Affiliate Professor, Prevention Research Center  
The Pennsylvania State University  
[rjf8@psu.edu](mailto:rjf8@psu.edu)

**DOI: 10.13140/RG.2.2.23767.06564**

# EARLY CHILDHOOD PROGRAM QUALITY IMPROVEMENT/INDICATOR MODEL (ECPQI2M4©) & DIFFERENTIAL MONITORING LOGIC MODEL AND ALGORITHM (DMLMA©) Update (Fiene, 12/12/15)

## Legend:

NRC = National Resource Center for Health and Safety in Child Care

AAP = American Academy of Pediatrics

APHA = American Public Health Association

OHS = Office of Head Start

ACF = Administration for Children and Families

OCC = Office of Child Care

ASPE = Assistant Secretary's Office for Planning and Evaluation

13I = *Thirteen Indicators of Quality Child Care*, ASPE

HSKI-C = *Head Start Key Indicators*

*Stepping Stones* = *Stepping Stones to Caring for Our Children*, NRC, AAP, APHA

PD = Professional Development, Training, Technical Assistance, Mentoring

PQ = Quality Rating and Improvement Systems (QRIS), Quality Improvements

TCO/TRC = Theory of Regulatory Compliance/Outcomes

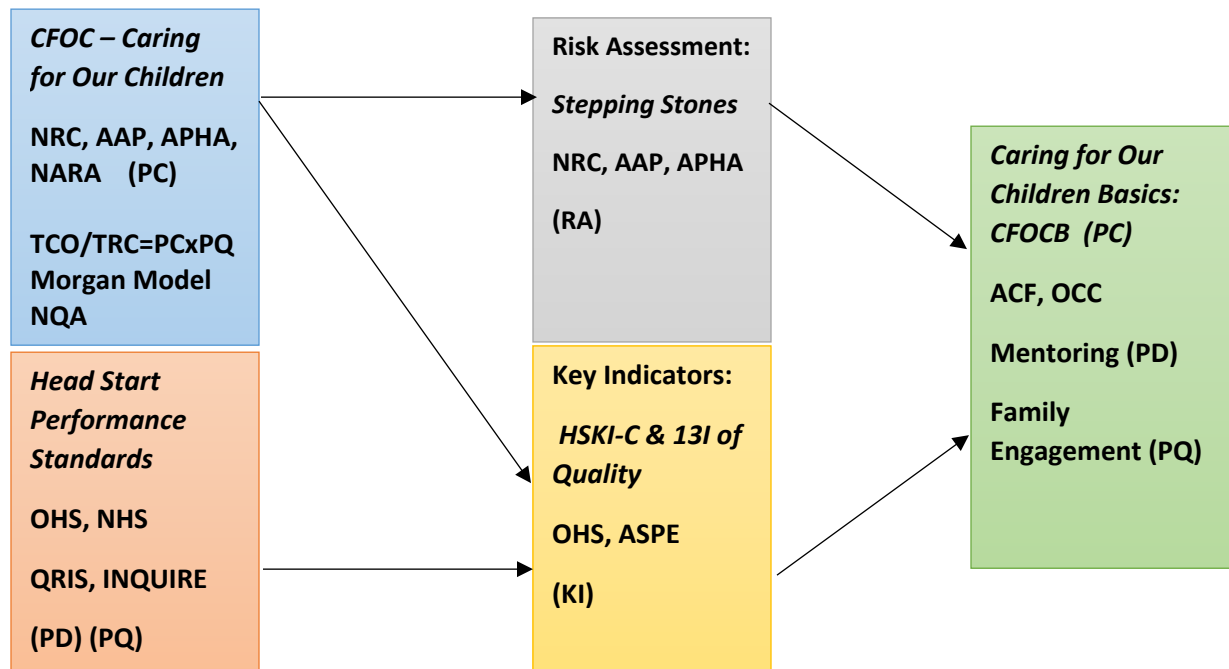
## Comprehensive Reviews

## Abbreviated Reviews

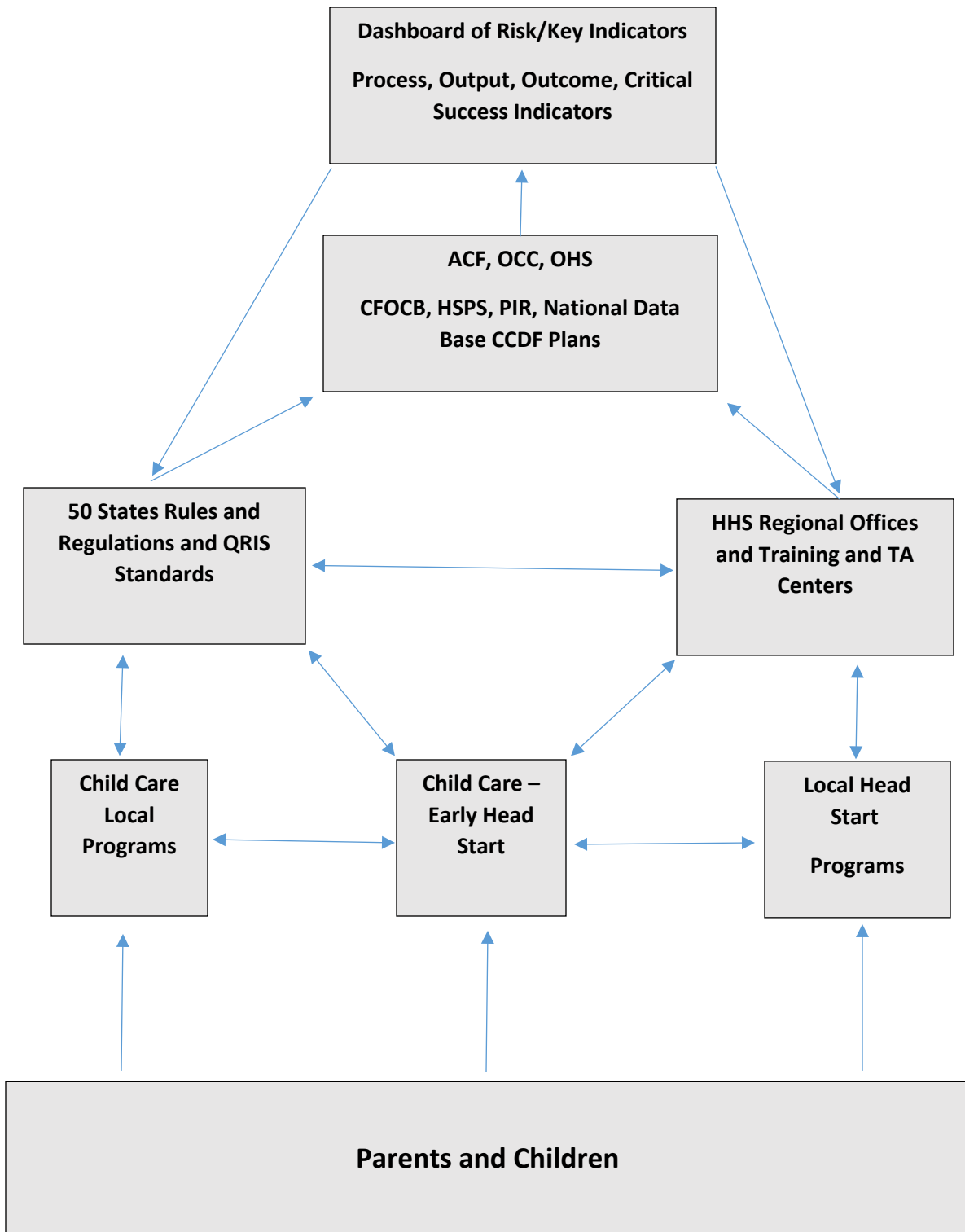
## Differential Monitoring

### *Absolute Paradigm*

### *Relative Paradigm*



## National Differential Monitoring Conceptual Framework (Fiene, 2016)





## National Differential Monitoring Conceptual Framework Brief Explanation:

The key elements for this conceptual framework is the emphasis on data utilization via key indicators and risk assessment which results in targeted/differential monitoring of programs via a state, regional, and national data base. Data would be collected at the local level in programs (child care (centers, homes, group homes); Head Start programs; child care/early Head Start programs, etc...) and would be monitored at the state and regional levels. The data via monitoring reports, CCDF plans, etc.. would move from the state and regional levels to the national level at ACF to form a national data base. From the national data base, a series of key indicator, risk assessment, process, output, outcome and critical success indicators would be culled (dashboard) from the full comprehensive data base to determine the levels of future reviews and monitoring of states and programs.

These indicators would be fed back to the regional offices and states with states being able to do the same with their respective licensing systems in reviews of local programs. The data from the comprehensive data base would also be fed back to the states, regional offices and the training & technical assistance offices to focus specific training and technical assistance based upon the results of the monitoring reviews. Within this conceptual framework, it is proposed to use a professional development passport within state professional development systems/registries which has badges attached for ongoing training & technical assistance for individual ECE staff. This professional development passport could provide the basis of a document (it would contain all the training received by the individual via a stamp/badge articulation documentation process) that would be transferable from state to state similar to how a regular passport is used as identification in moving from country to country. This could potentially become a national credentialing/licensing system for ECE staff.

This conceptual framework would take into account the collecting and analyzing of data and its subsequent utilization for training & technical assistance. All the components/key elements for such a system have been set up by ACF, now what we need to do is put all the pieces together into a unified monitoring system.

## Theory of Regulatory Compliance Algorithm (2/17)

- 1)  $\Sigma R = C$
- 2) Review C history x 3 yrs
- 3)  $NC + C = CI$
- 4) If  $CI = 100 \rightarrow KI$
- 5) If  $KI > 0 \rightarrow CI$  or if  $C < 100 \rightarrow CI$
- 6) If  $RA (NC\% > 0) \rightarrow CI$
- 7)  $KI + RA = DM$
- 8)  $KI = ((A)(D)) - ((B)(E)) / \text{sqrt}((W)(X)(Y)(Z))$
- 9)  $RA = \Sigma R1 + \Sigma R2 + \Sigma R3 + \dots \Sigma Rn / N$
- 10)  $(TRC = 99\%) + (\phi = 100\%)$
- 11)  $(CI < 100) + (CIPQ = 100) \rightarrow KI (10\% CI) + RA (10-20\% CI) + KIQP (5-10\% \text{ of } CIPQ) \rightarrow OU$

### Legend:

R = Rules/Regulations/Standards

C = Compliance with Rules/Regulations/Standards

NC = Non-Compliance with Rules/Regulations/Standards

CI = Comprehensive Instrument for determining Compliance

$\phi$  = Null

KI = Key Indicators

$KI \geq .26$  Include

$KI \leq .25$  Null, do not include

RA = Risk Assessment

$\Sigma R1$  = Specific Rule on Likert Risk Assessment Scale (1-8; 1 = low risk, 8 = high risk)

N = Number of Stakeholders

DM = Differential Monitoring

TRC = Theory of Regulatory Compliance

CIPQ = Comprehensive Instrument Program Quality

KIPQ = Key Indicators Program Quality

OU = Outcomes

A = High Group + Programs in Compliance on Specific Compliance Measure (R1...Rn).

B = High Group + Programs out of Compliance on Specific Compliance Measure (R1...Rn).

E = Low Group + Programs in Compliance on Specific Compliance Measure (R1...Rn).

D = Low Group + Programs out of Compliance on Specific Compliance Measure (R1...Rn).

W = Total Number of Programs in Compliance on Specific Compliance Measure (R1...Rn).

X = Total Number of Programs out of Compliance on Specific Compliance Measure (R1...Rn).

Y = Total Number of Programs in High Group ( $\Sigma R = 98+$ ).

Z = Total Number of Programs in Low Group ( $\Sigma R \leq 97$ ).

High Group = Top 25% of Programs in Compliance with all Compliance Measures ( $\Sigma R$ ).

Low Group = Bottom 25% of Programs in Compliance with all Compliance Measures ( $\Sigma R$ ).

## Regulatory Compliance Matrices

### 2 x 2 Matrix (In vs Out of compliance x High vs Low Groups):

<b>A</b>	<b>B</b>
<b>C</b>	<b>D</b>

(A = In compliance + High Group)(B = In compliance + Low Group)(C = Out of Compliance + High Group)(D = Out of Compliance + Low Group); **B = false positives**; **C = false negatives**; **A + D > B + C**; **B > C**; **A + D = + results**.

### 2 x 3 Matrix (In vs Out of compliance x 100% vs Substantial vs Low Compliance Groups):

<b>A</b>	<b>B</b>	<b>C</b>
<b>D</b>	<b>E</b>	<b>F</b>

(A = In compliance + 100% Group)(B = In compliance + Substantial Compliance Group)(C = In compliance + Low Group)(D = Out of compliance + 100% Group)(E = Out of compliance + Substantial Compliance Group)(F = Out of compliance + Low Group); **C = false positives**; **D, E = false negatives**; **B > A > C**; **B + F = + results**.

### 3 x 2 Matrix (In vs Partial vs Out of compliance x High vs Low Groups):

<b>A</b>	<b>B</b>
<b>C</b>	<b>D</b>
<b>E</b>	<b>F</b>

(A = In compliance + High Group)(B = In compliance + Low Group)(C = Partial compliance + High Group)(D = Partial compliance + Low Group)(E = Out of compliance + High Group)(F = Out of compliance + Low Group); **B = false positives**; **E = false negatives**; **A > C > B > D**; **A + F = +results**.

### 3 x 3 Matrix (In vs Partial vs Out of compliance x 100% vs Substantial vs Low Compliance Groups):

<b>A</b>	<b>B</b>	<b>C</b>
<b>D</b>	<b>E</b>	<b>F</b>
<b>G</b>	<b>H</b>	<b>I</b>

(A = In compliance + 100% Group)(B = In compliance + Substantial Compliance Group)(C = In compliance + Low Group)(D = Partial compliance + 100% Group)(E = Partial compliance + Substantial Compliance Group)(F = Partial compliance + Low Group)(G = Out of compliance + 100% Group)(H = Out of compliance + Substantial Compliance Group)(I = Out of compliance + Low Group); **C = false positives**; **G, H = false negatives**; **B > A > D > E > C > F**; **B + D + I = + results**.

## **Theory of Regulatory Compliance and Regulatory Compliance Monitoring Paradigm Matrix Notes (Fiene, 2-12-17)**

### **Outline:**

- 2x2 absolute vs 3x3+ relative matrices.
- 2x2 In or Out x 100% or 0%.
- 3x3 100%, Substantial, Low x In, Partial, Out.
- TRC proposes 3x2 = 100%, Substantial, Low x In, Out.
- KI 2x2 or 3x2; RA 3x3 matrices.
- Normally distributed curve 3x3+ vs Skewed data 2x2 - visualize a normally distributed curve over the cells vs a very skewed curve over the 2 cells.
- ERS as 7x7 potential matrix.
- Use these matrices to explain RCMP and potential data analyses.
- Better analytical techniques for analyzing these matrices.
- Problem with 2x2 are the false negatives.
- Does a 3x3+ reduce the false negatives. Key question.
- What I have found over my 40+years is that I have as many questions as I have answers at this point, not sure that 2x2 or 3x2 are best matrices. What happens if we expand to a 7x7 matrix.
- Phi to Chi-square as the preferred statistic?
- Would Matrix Algebra be more appropriate.
- First time tying KI and RA together via 2x2 and 3x3 matrices. Common analytical framework.

### **Research Questions:**

What are the differences between a 2x2 vs 2x3 vs 3x3 matrices? This will account for absolute, relative and substantial compliance ranges.

What is the impact of having 2x2, 2x3, and 3x3 on false negatives?

What are the results with 100% vs 99-98% and low compliance groups?

What are the differences between samples and full data sets?

Relationship between PC and PQ? Linear or non-linear

### Matrices:

<b>A</b>	<b>B</b>
<b>C</b>	<b>D</b>

2 x 2 = I/O x H/L (I = In compliance)(O = Out of compliance)(H = High Group)(L = Low Group)

**A + D = positive+ results, to be expected**

B = false positives

**C = false negatives**

A + D > B + C

B > C

Class ARC Matrix

<b>A</b>	<b>B</b>	<b>C</b>
<b>D</b>	E	<b>F</b>

3 x 2 = H/S/L x I/O (S = Substantial Compliance) or 3 x 3 with I/P/O where P = Partial.

A = 100% compliance

B = Substantial compliance

C = Low compliance

C = false positives

**D = false negatives**

B > A > C

**B + F = + results, to be expected**

Fiene TRC Matrix

<b>A</b>	<b>B</b>	<b>C</b>
<b>D</b>	<b>E</b>	<b>F</b>
<b>G</b>	<b>H</b>	<b>I</b>

3 X 3+ = H/M/L x H/M/L

A = Low probability + low risk

E= Medium probability + medium risk

I= High probability + high risk

**A > B > C > D > E > F > G > H > I**

Fiene RA Matrix

**Classification Matrix & Sensitivity Analysis for Validating Licensing Key indicator Systems**  
**Technical Research Note (Fiene, 2017)**

	1	2	3	5	7	8	10	Comments
<b>A</b>	1.00	1.00	1.00	0.00	0.00	1.00	1.00	Perfect
<b>B</b>	0.52	0.52	0.52	0.48	0.48	0.52	0.04	Random
<b>C</b>	0.71	0.96	0.94	0.04	0.29	0.84	0.70	False (-)
<b>D</b>	0.94	0.78	0.71	0.22	0.06	0.81	0.70	False (+)
<b>E</b>	-----	0.00	0.00	1.00	-----	0.00	-----	False +100%
<b>F</b>	0.00	0.00	0.00	1.00	1.00	0.00	-1.00	False+-100
<b>H</b>	0.45	0.46	0.40	0.54	0.55	0.46	-0.08	Random

**Measures:**

1 = Sensitivity	$TPR = TP / (TP + FN)$
2 = Specificity	$SPC = TN / (FP + TN)$
3 = Precision	$PPV = TP / (TP + FP)$
5 = False Positive	$FPR = FP / (FP + TN)$
7 = False Negative	$FNR = FN / (FN + TP)$
8 = Accuracy	$ACC = (TP + TN) / (P + N)$
10 = Correlation	$((TP)(TN)) - ((FP)(FN)) / \sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}$

PP = Predicted Positive = CI+

PN = Predicted Negative = CI-

TP= True Positive = KI+

TN = True Negative =KI-

	TRUE POSITIVE (TP)(KI+)	TRUE NEGATIVE (TN)(KI-)
PREDICTED POSITIVE (PP)(CI+)	++	+-
PREDICTED NEGATIVE (PN)(CI-)	-+	--

**CI+/CI-/KI+/KI-**

**A = 25/0/0/25 - Perfect match between CI and KI.**

**B = 13/12/12/13 - Random matching between CI and KI.**

**C = 17/7/1/25 - KI+ x CI- (False-)**

**D = 17/1/7/25 - KI- x CI+ (False+)**

**E = 0/0/50/0 - KI- x CI+ unlikely**

**F = 0/25/25/0 - False + & - 100% unlikely**

**H = 20/24/30/26 - Random matching between CI and KI.**

## Technical Detail Notes: Validation Updates to the Fiene Key Indicator Systems

January 2015

These notes will provide guidance on validating existing Key Indicator Licensing Systems. These notes are based upon the last three years of research and data analysis in determining the best means for conducting these validation studies.

These notes are based upon existing Key Indicator Systems in which data can be drawn from an already present data base which contains the comprehensive instrument (total compliance data) and the key indicator instrument (key indicator rule data). When this is in place and it can be determined how licensing decisions are made: full compliance with all rules or substantial compliance with all rules to receive a license, then the following matrix can be used to begin the analyses (see Figure 1):

<b>Figure 1</b>	<i>Providers who fail the Key Indicator review</i>	<i>Providers who pass the Key Indicator review</i>	<i>Row Totals</i>
<i>Providers who fail the Comprehensive review</i>	W	X	
<i>Providers who pass the Comprehensive Review</i>	Y	Z	
<i>Column Totals</i>			<i>Grand Total</i>

A couple of annotations regarding Figure 1.

**W + Z** = the number of agreements in which the provider passed the Key Indicator review and also passed the Comprehensive review.

**X** = the number of providers who passed the Key Indicator review but failed the Comprehensive review. This is something that should not happen, but there is always the possibility this could occur because the Key Indicator Methodology is based on statistical methods and probabilities. We will call these False Negatives (FN).

**Y** = the number of providers who failed the Key Indicator review but passed the Comprehensive review. Again, this can happen but is not as much of a concern as with “**X**”. We will call these False Positives (FP).

Figure 2 provides an example with actual data from a national organization that utilizes a Key Indicator System. It is taken from 50 of its program providers.

<b>Figure 2</b>	<i>Providers who fail the Key Indicator review</i>	<i>Providers who pass the Key Indicator review</i>	<i>Row Total</i>
<i>Providers who fail the Comprehensive review</i>	25	1	26
<i>Providers who pass the Comprehensive Review</i>	7	17	24
<i>Column Total</i>	32	18	50

To determine the agreement ratio, we use the following formula:

$$\frac{A}{A + D}$$

Where **A** = Agreements and **D** = Disagreements.



Based upon Figure 2,  $A + D = 42$  which is the number of agreements; while the number of disagreements is represented by  $B = 1$  and  $C = 7$  for a total of 8 disagreements. Putting the numbers into the above formula:

$$\frac{42}{42 + 8}$$

Or

$$.84 = \text{Agreement Ratio}$$

The False Positives (FP) ratio is .14 and the False Negatives (FN) ratio is .02. Once we have all the ratios we can use the ranges in Figure 3 to determine if we can validate the Key Indicator System. The FP ratio is not used in Figure 3 but is part of the Agreement Ratio.

**Figure 3 – Thresholds for Validating the Fiene Key Indicators for Licensing Rules**

<u>Agreement Ratio Range</u>	<u>False Negative Range</u>	<u>Decision</u>
(1.00) – (.90)	.05+	Validated
(.89) – (.85)	.10 - .06	Borderline
(.84) – (.00)	.11 or more	Not Validated

---

### RESOURCES AND NOTES

For those readers who are interested in finding out more about the Key Indicator Methodology and the more recent technical updates as applied in this paper in actual state examples, please see the following publication:

Fiene (2014). *ECPQIM4©: Early Childhood Program Quality Indicator Model4*, Middletown: PA; Research Institute for Key Indicators LLC (RIKI). (<http://drfiene.wordpress.com/riki-reports-dmlma-ecpqim4/>)

In this book of readings/presentations are examples and information about differential monitoring, risk assessment, key indicators, validation, measurement, statistical dichotomization of data, and regulatory paradigms. This publication delineates the research projects, studies, presentations, & reports completed during 2013-14 in which these updates are drawn from.

For those readers interested in a historical perspective to the development of the Key Indicator methodology and licensing measurement, please see the following publications (most of these publications are available at the following website (<http://rikoinstitute.wikispaces.com/home>):

***For additional information regarding this paper please contact:***

Dr Richard Fiene

Research Institute for Key Indicators LLC (RIKI)

41 Grandview Avenue

Middletown, PA. 17057

717-944-5868

<http://DrFiene.wordpress.com/home>

The logo for RIKI LLC, featuring the letters "RIKI" in a large, bold, serif font, with "LLC" in a smaller, bold, sans-serif font to its right. The text is white and set against a dark, rectangular background with a subtle gradient.

KEY INDICATOR TECHNICAL NOTES (12-8-15) RJF (this note updates a previous technical note from earlier in 2015 regarding this same topic):

Each state/jurisdiction will be different when applying the Key Indicator Methodology but there are some guiding principles that should be used:

- 1) Sample size should be around 100-200 programs. Less than 100 may not produce significant results and indicators will be missed. Over 200 programs will provide too many indicators reaching significance.
- 2) Set the p value to .01 ( $p < .01$ ).  $P < .05$  is too lenient and  $p < .001$  is too stringent.  $P < .01$  gives a proper balance for the number of indicators a state/jurisdiction will need.
- 3) The best model to use is the 100% for the high group (100-99% can also be used) with the middle programs not being used and the bottom 25% being used for the low group. The worse model to use is 100% as the high group and 99% or less as the low group. Too much error variance in the programs is introduced with an increase in making false negatives and the phi and Pearson correlations drop off significantly.
- 4) Select a moderate number of key indicators, don't select too few. It is more reliable to go with a few additional indicators than using too few.
- 5) Minimize false negatives by using the model described in #3 above.

## Validation of the Key Indicator Methodology: Two Examples

Richard Fiene, Ph.D.

June 2015

### Introduction

The purpose of this paper is to address the validation of the key indicator methodology as suggested in the *ASPE White Paper on ECE Monitoring* (2015). It was so accurately pointed out in this *White Paper* regarding the need to continue to access and validate differential monitoring which generally consists of the key indicator and risk assessment methods.

Over the past 35 years various aspects of differential monitoring have been assessed and validated. For example, studies by Kontos and Fiene (1987) and Fiene (2000) demonstrated the relationship between key indicators and child development outcomes. In 2002, another *ASPE White Paper on the Thirteen Indicators of Quality Child Care: A Research Update* summarized the research over the previous 20 years in demonstrating a core set of key indicator risk assessment standards. More recently, a study completed in Georgia (Fiene, 2014) validated the use of core rules in a risk assessment and differential monitoring approach. And in 2012, a study was done in California which demonstrated the time savings in using a key indicator approach. And finally, in 2013-14, a study was done in the national Head Start program in which their key indicator approach (Head Start Key Indicators (HSKI)) validated the decision making ability of key indicators in which an 84% - 91% agreement was found between the HSKI and Full Compliance Reviews. The focus of this paper will be on the latest findings from Head Start since these findings have not been published to date.

The National Child Care Licensing Study (2011) and the National Center for Child Care Quality Improvement (2014) have reported the significant use of differential monitoring, key indicators and risk assessment methods by many states throughout the country. And with the reauthorization of CCDBG (2014) and the increased emphasis on ECE program monitoring there is an increased need to validate these approaches. This paper is the beginning attempt to begin this process focusing on the key indicator method.

## Methodology

This validation method is based upon existing Key Indicator Systems in which data can be drawn from an already present data base which contains the comprehensive instrument (total compliance data) and the key indicator instrument (key indicator rule data). When this is in place and it can be determined how licensing decisions are made: full compliance with all rules or substantial compliance with all rules to receive a license, then the following matrix can be used to begin the analyses (see Figure 1):

<b>Figure 1</b>	<i>Providers who fail the Key Indicator review</i>	<i>Providers who pass the Key Indicator review</i>	<i>Row Totals</i>
<i>Providers who fail the Comprehensive review</i>	<b>W</b>	<b>X</b>	
<i>Providers who pass the Comprehensive Review</i>	<b>Y</b>	<b>Z</b>	
<i>Column Totals</i>			<i>Grand Total</i>

A couple of annotations regarding Figure 1.

**W + Z** = the number of agreements in which the provider passed the Key Indicator review and also passed the Comprehensive review.

**X** = the number of providers who passed the Key Indicator review but failed the Comprehensive review. This is something that should not happen, but there is always the possibility this could occur because the Key Indicator Methodology is based on statistical methods and probabilities. We will call these False Negatives (FN).

**Y** = the number of providers who failed the Key Indicator review but passed the Comprehensive review. Again, this can happen but is not as much of a concern as with “**X**”. We will call these False Positives (FP).

Figure 2 provides an example with actual data from a national organization that utilizes a Key Indicator System. It is taken from 50 of its program providers.

<b>Figure 2</b>	<i>Providers who fail the Key Indicator review</i>	<i>Providers who pass the Key Indicator review</i>	<i>Row Total</i>
<i>Providers who fail the Comprehensive review</i>	25	1	26
<i>Providers who pass the Comprehensive Review</i>	7	17	24
<i>Column Total</i>	32	18	50

To determine the agreement ratio, we use the following formula:

$$\frac{A}{A + D}$$

Where A = Agreements and D = Disagreements.

Based upon Figure 2, A + D = 42 which is the number of agreements; while the number of disagreements is represented by B = 1 and C = 7 for a total of 8 disagreements. Putting the numbers into the above formula:

$$\frac{42}{42 + 8}$$

Or

$$.84 = \text{Agreement Ratio}$$

The False Positives (FP) ratio is .14 and the False Negatives (FN) ratio is .02. Once we have all the ratios we can use the ranges in Figure 3 to determine if we can validate the Key Indicator System. The FP ratio is not used in Figure 3 but is part of the Agreement Ratio.

**Figure 3 – Thresholds for Validating the Fiene Key Indicators for Licensing Rules**

<u>Agreement Ratio Range</u>	<u>False Negative Range</u>	<u>Decision</u>
(1.00) – (.90)	.05+	Validated
(.89) – (.85)	.10 - .06	Borderline
(.84) – (.00)	.11 or more	Not Validated

## Results

The following results are from a study completed in 2014 using Head Start data where HSKI reviews were compared with comprehensive reviews to make certain that additional non-compliance was not found when HSKI tools were administered to programs.

There was an 84% - 91% (see Table 1) agreement between the HSKI and Comprehensive Reviews which would indicate that the HSKI method was validated in Head Start based upon Figure 3 above in the Methodology section.

**FY 2015 HSKI Agreement Table 1**

FY 2015 HSKI Agreement Tables with Combined OHSMS Data from FYs 2012, 2013, and 2014

	% agreement	Sensitivity
<b>FIS</b>	91%	63%
<b>GOV/SYS</b>	84%	63%
<b>SR</b>	87%	52%

### **Fiscal (5)**

- FIS1.1 - Effective financial management systems (D, I, T)
- FIS2.1 - Timely and complete financial records (D)
- FIS4.1 - Signed and approved time records (T)
- FIS5.3 - NFS contributions are necessary and reasonable (D)
- FIS6.2 - Complete and accurate equipment records (D, T)

### **SR (9)**

- CDE1.2 - System to track, use, and report on SR goals (I)
- CDE2.1 - Evidenced-based curriculum (I)
- CDE3.1 – Individualizing (I)
- CDE3.4 - Child access to mental health services (I)
- CDE4.1 - Teacher qualifications (S)
- CHS1.5 - Health services tracking system (I)
- CHS2.2 - Referrals for children with disabilities to LEA or Part C Agency
- FCE2.3 - Access to mental health services (I)
- FCE5.3 - Coordination with LEAs and Part C Agencies

### **GOV/SYS (9)**

- GOV2.1 - Training and Technical Assistance for GB and PC (I)
- GOV2.2 - GB responsibilities regarding program administration and operations (I)
- GOV3.1 - Reporting to GB and PC (I)
- GOV2.4 - PC submits program activity decisions to GB (I)
- SYS1.2 - Annual Self-Assessment (I)

- SYS4.1 - Communication mechanisms (I)
- SYS5.2 - Publication and availability of an Annual Report (I)
- SYS2.1 - Ongoing Monitoring (I)
- SYS5.1 - Record-keeping (I)

I = Interview

D = Document Review

T = Transaction Review

S = Staff files

## **Discussion**

This paper presents a validation methodology to validate the differential monitoring approach that utilizes key indicators. This is an area that needs additional research as many more states began to think about employing the various approaches for differential monitoring involving risk assessment and key indicators.

The results from this paper are very encouraging in that they clearly demonstrate that a very large delivery system, the national Head Start program, can utilize key indicators (HSKI – Head Start Key Indicators) for a differential monitoring approach (Aligned Monitoring System).

## **For additional information regarding this paper:**

*Richard Fiene, Ph.D.*

*Research Institute for Key Indicators (RIKI)*

<http://RIKInstitute.wikispaces.com>

[RIKI.Institute@gmail.com](mailto:RIKI.Institute@gmail.com)



## Appendix

A more recent validation study has been completed in the Province of Ontario, Canada where they compared three sets of Key Indicators over three calendar years in a similar fashion to the Head Start study reported above. Below are the results of these analyses.

**Validation Summary**

Year	Key Indicators	Agreement Ratio
2014	29 Indicators	0.90
	35 Indicators	0.92
	41 Indicators	0.94
2013	29 Indicators	0.90
	35 Indicators	0.92
	41 Indicators	0.93
2012	29 Indicators	0.91
	35 Indicators	0.93
	41 Indicators	0.94

*Note. The key indicators are validated when the agreement ratio is 0.90 or above.*

## **Technical Detail Updates to the Fiene Key Indicator Methodology**

**January 2015**

The Key Indicator Methodology has recently been highlighted in a very significant Federal Office of Child Care publication series on Contemporary Licensing Highlights. In that Brief the Key Indicator Methodology is described as part of a differential monitoring approach along with the risk assessment methodology. Because of the potential increased interest in the Key Indicator Methodology, a brief update regarding the technical details of the methodology is warranted. For those readers who are interested in the historical development of Key Indicators I would suggest they download the resources available at the end of the paper.

This brief paper provides the technical and statistical updates for the key indicator methodology based upon the latest research in the field related to licensing and quality rating & improvement systems (QRIS). The examples will be drawn from the licensing research but all the reader needs to do is substitute “rule” for “standard” and the methodology holds for QRIS.

Before proceeding with the technical updates, let me review the purpose and conceptual underpinning of the Key Indicator Methodology. Key Indicators generated from the methodology are not the rules that have the highest levels of non-compliance nor are they the rules that place children most at risk of mortality or morbidity. Key Indicators are generally somewhere in the middle of the pack when it comes to non-compliance and risk assessment. The other important conceptual difference between Key Indicators and risk assessment is that only Key Indicators statistically predict or are predictor rules of overall compliance with all the rules for a particular service type. Risk assessment rules do not predict anything other than a group of experts has rated these rules as high risk for children’s mortality/morbidity if not complied with.

Something that both Key Indicators and risk assessment have in common is through their use one will save time in their monitoring reviews because you will be looking at substantially fewer rules. But it is only with Key Indicators that you can statistically predict additional compliance or non-compliance; this is not the case with risk assessment in which one is only looking at those rules which are a state’s high risk rules. And this is where differential monitoring comes into play by determining which programs are entitled to either Key Indicators and/or risk assessment for more abbreviated monitoring reviews rather than full licensing reviews (the interested reader

should see the *Contemporary Licensing Series on Differential Monitoring, Risk Assessment and Key Indicators* published by the Office of Child Care.

## Technical and Statistical Framework

One of the first steps in the Key Indicator Methodology is to sort the licensing data into high and low groups, generally the highest and lowest licensing compliance with all the rules can be used for this sorting. 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 child care rule (see Figure 1). In some cases, especially where there is very high compliance with the rules and the data are extremely skewed, it may be necessary to use all those programs that are in full (100%) compliance with all the rules as the high group. The next step is to look at each rule and determine if it is in compliance or out of compliance with the rule. This result is cross-referenced with the High Group and the Low Group as depicted in Figure 1.

<b>Figure 1</b>	<i>Providers In Compliance on Rule</i>	<i>Programs Out Of Compliance on Rule</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>

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

**Figure 2 – Formula for Fiene Key Indicator Coefficient**

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

**Figure 3 – Legend for the Cells within the Fiene Key Indicator Coefficient**

*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.*

Once the data are run through the formula in Figure 2, the following chart (Figure 4) can be used to make the final determination of including or not including the rule as a key indicator. Based upon the chart in Figure 4, it is best to have a Key Indicator Coefficient approaching +1.00 however that is rarely attained with licensing data but has occurred in more normally distributed data.

Continuing with the chart in Figure 4, if the Key Indicator Coefficient is between +.25 and -.25, this indicates that the indicator rule is unpredictable in being able to predict overall compliance with the full set of rules. Either a false positive in which the indicator appears too often in the low group as being in compliance, or a false negative in which the indicator appears too often in the high group as being out of compliance. This can occur with Key Indicator Coefficients above +.25 but it becomes unlikely as we approach +1.00 although there is always the possibility that other rules could be found out of compliance. Another solution is to increase the number of key indicator rules 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 Key Indicator Coefficient is if it is between -.26 and -1.00, this indicates that the indicator is a terrible predictor because it is doing just the opposite of the decision we want to make. The indicator rule would predominantly be in compliance with the low group rather than the high group so it would be statistically predicting overall non-compliance. This is obviously something we do not want to occur.

Figure 5 gives the results and decisions for a QRIS system. The thresholds in a QRIS system are increased dramatically because QRIS standard data are less skewed than licensing data and a

more stringent criterion needs to be applied in order to include particular standards as Key Indicators.

**Figure 4 – Thresholds for the Fiene Key Indicators for Licensing Rules**

<u>Key Indicator Range</u>	<u>Characteristic of Indicator</u>	<u>Decision</u>
(+1.00) – (+.26)	Good Predictor	Include
(+.25) – (-.25)	Unpredictable	Do not Include
(-.26) – (-1.00)	Terrible Predictor	Do not Include

**Figure 5 – Thresholds for the Fiene Key Indicators for QRIS Standards**

<u>Key Indicator Range</u>	<u>Characteristic of Indicator</u>	<u>Decision</u>
(+1.00) – (+.76)	Good Predictor	Include
(+.75) – (-.25)	Unpredictable	Do not Include
(-.26) – (-1.00)	Terrible Predictor	Do not Include

## RESOURCES AND NOTES

For those readers who are interested in finding out more about the Key Indicator Methodology and the more recent technical updates as applied in this paper in actual state examples, please see the following publication:

Fiene (2014). *ECPQIM4©: Early Childhood Program Quality Indicator Model4*, Middletown: PA; Research Institute for Key Indicators LLC (RIKI). (<http://drfiene.wordpress.com/riki-reports-dmlma-ecpqim4/>)

In this book of readings/presentations are examples and information about differential monitoring, risk assessment, key indicators, validation, measurement, statistical dichotomization of data, and regulatory paradigms. This publication delineates the research projects, studies, presentations, & reports completed during 2013-14 in which these updates are drawn from.

For those readers interested in a historical perspective to the development of the Key Indicator methodology and licensing measurement, please see the following publications (most of these publications are available at the following website (<http://rikoinstitute.wikispaces.com/home>):

Lahti, Elicker, Zellman, & Fiene (2014). Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs, *Early Childhood Research Quarterly*, available online 9 June 2014, doi:10.1016/j.ecresq.2014.04.005.

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.

Zellman & Fiene (2012). *Validation of quality rating and improvement systems for early care and education and school-age care*, Washington, D.C.: OPRE and Child Trends.

Fiene & Carl (2011). Child Care Quality Indicators Scale, in T Halle (Ed.), *Quality Rating and Improvement Systems Tool Kit*, Washington, D.C.: Child Trends.

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, 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, & Kroh (2000). Measurement tools and systems, in *Licensing Curriculum*, National Association for Regulatory Administration, Minneapolis, Minnesota.

Fiene (1997). Potential solution to the child day care trilemma related to quality, accessibility and affordability. *Child Care Information Exchange*, September, 57-60.

Fiene (1997). Human services licensing information system. *National Association for Regulatory Administration: Research Column*, Spring, 9-10.

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.

Kuhns & Fiene (1995). Promoting health and safety in child care programs, *Child Care Bulletin*, January-February (1), 3.

Fiene (1995). *National early childhood program accreditation standards*. Atlanta, Georgia: National Early Childhood Program Accreditation Commission.

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.

Fiene (1994). The case for national early care and education standards: Key indicator/predictor state child care regulations, *National Association of Regulatory Administration*, summer 1994, 6-8.

Fiene (1991). New early childhood research, evaluation and training program has impact on Pennsylvania for the 1990's, *Dimensions*, Fall, 4.

Fiene (1988). Human services instrument based program monitoring and indicator systems, in *Information Technology and the Human Services*, B. Glastonburg, W. LaMendola, & S. Toole, editors, Chichester, England: John Wiley and Sons.

Fiene & McDonald (1987). *Instrument based program monitoring for child welfare*, Portland, Maine: University of Southern Maine.

Fiene (1987). Using licensing data in human service programs, in *Licensing*, H. Hornby, editor, Portland, Maine: University of Southern Maine.

Fiene (1987). The indicator system, in *Evaluation and outcome monitoring*, H. Hornby, editor, Portland, Maine: University of Southern Maine.

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.

Fiene (1987). Indicator checklist system, in *Maximizing the Use of Existing Data Systems*, Portland, Maine: University of Southern Maine.

Fiene (1986). State child care regulatory, monitoring and evaluation systems as a means for ensuring quality child development programs, in *Licensing of Children's Services Programs*, Richmond, Virginia: Virginia Commonwealth University School of Social Work. (ERIC/ECE ED322997)

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.

Kontos & Fiene (1986). Predictors of quality and children's development in day care, in *Licensing of Children's Services Programs*, Richmond, Virginia: Virginia Commonwealth University School of Social Work.

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

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

Fiene & Nixon (1983). *Indicator checklist system for day care monitoring*, Washington, D.C.: National Children's Services Monitoring Consortium.

Fiene & Nixon (1981). *An instrument based program monitoring information system: A new tool for day care monitoring*, Washington, D.C.: National Children's Services Monitoring Consortium.

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

Fiene, Cardiff, & Littles (1975). Ecological monitoring information system, *In the Best Interests of Children*, July-September, 1975.

*For additional information regarding this paper please contact:*

Dr Richard Fiene

Research Institute for Key Indicators LLC (RIKI)

41 Grandview Avenue

Middletown, PA. 17057

717-944-5868

<http://DrFiene.wordpress.com/home>





## **Key Indicator Methodology Technical Note(2): The Dichotomization and Bi-Polarization of the Matrix Data Base**

**Richard Fiene, Ph.D.**

**June 2015**

This latest technical note updates the thresholds for the high and low groups within the key indicator matrix. This technical note is based upon the latest studies during the early 2015 time frame in which very large data distributions were available to test certain criteria with the key indicator methodology. Because of the extreme skewness present in licensing/regulatory data, certain statistical adjustments need to be made so that the analyses performed reflect the distribution of data. One of these statistical adjustments is the dichotomization of data which is generally not suggested with the exception of very skewed data. Since licensing data are so skewed, this adjustment has been used throughout the key indicator methodology. However, an additional adjustment is now warranted given not only the skewness of data but also because of the data being nominal in nature. This additional adjustment I am calling the bi-polarization of data in order to accentuate the differences between the high and low groups within the key indicator matrix.

I have tested several data sets utilizing bi-polarization and found that the results are more significant with its use than without its use. Please keep in mind that licensing data is very different from other forms of data found in the early care and education (ECE) research literature. It is not like the ERS or CLASS data which is more normally distributed and lends itself to more parametric statistical analyses. Licensing data are nominal in nature and always very skewed which means that more non-parametric methods are warranted, such as phi coefficient and dichotomization of data. An example of how this actually works may help.

Licensing data are measured as either being in or out of compliance. There is no middle ground, it is not measured on a Likert scale. Therefore it is nominal in nature, either it is all there or it is not. Licensing data are also measured in the sense that all rules are created equally, in other words, they all have the same weight or importance, such as 1 = compliance; 0 = non-compliance. Being in full 100% compliance which means 0 violations is the goal of a regulatory/licensing system. One does not want to see many violations of the rules because this will place children at risk of harm and the purpose of an early care and education (ECE) licensing/regulatory system is to reduce the potential harm to children. In the licensing measurement literature, this 100% compliant group is generally labeled or considered the high

compliant group. With some licensing laws which allow substantial but not full 100% compliance with the full set of rules, it would then be allowable to have possibly 1 or 2 violations and still be considered in this high compliant group. The low compliant group has been generally any program that had any non-compliance or had 2 or more violations. When these two groups were compared to each individual rule utilizing the phi coefficient formula it was found that a more accurate approach was to accentuate or increase the difference between the high and low groups by eliminating the intervening violations in following manner: high group of 0 violations; 1-4 violations being eliminated; 5+ violations defined as the low group. This additional bi-polarization of data helped to accentuate the differences in calculating the phi coefficient and provided a more sensitive key indicator tool.

Another data distribution issue that should be addressed here that justifies the above cutoffs is that there is very little variance in licensing/regulatory data. Generally the frequency distribution is 20 or less and the average set of rules is over 200 rules. So the frequency distribution is extremely skewed within less than 10% of the potential data distribution. Also, the majority of programs are 100% in compliance with all the rules. And an additional complication is that the scoring of each rule is scored as if it had an equal risk value when in reality the rules can place children at either great risk to relatively little risk if found non-compliant. These measurement issues are very different than in other measurement systems such as ERS or CLASS. The important message to take from this is that rules are not a ruler, they do not measure things equally and cannot be analyzed or compared to other measurement systems that are more normally distributed.

Although licensing is part of the program quality continuum in establishing basic health and safety standards for children, it is a system with measurement limitations that can only be compared on a nominal basis making several statistical adjustments as suggested above necessary.



DR RICHARD FIENE, RESEARCH PSYCHOLOGIST  
RESEARCH INSTITUTE FOR KEY INDICATORS, LLC (RIKI)  
1871 GRAMERCY PLACE  
HUMMELSTOWN, PENNSYLVANIA 17036  
MOBILE 717-598-8908  
FIENE@RIKINSTITUTE.COM  
RIKINSTITUTE.COM

PA ENTITY #: 4193691 D-U-N-S #: 011323030

## **Problem Solving Coaching = Online Pinging: How to Make Coaching Both Effective and Efficient and Some Additional Individual Learning Advantages**

**Richard Fiene, PhD & Benjamin Levi, MD, PhD  
Penn State Prevention Research Center & College of Medicine**

**September 2017**

The purpose of this short paper is to introduce a potentially new technology that can impact the professional development field as well as learning in general. It is presented for its heuristic value to get us thinking about the possibilities of this new technology as a new online delivery system.

We know that problem solving coaching is an effective quality improvement/professional development intervention (Training, Technical Assistance, and Quality Rating and Improvement Systems) but one that is not particularly efficient. It is very time intensive which drives up cost but it is so much more effective than run of the mill professional development interventions that revolve around workshop or lecture type delivery. (***Mathematical Policy Research (2011) has completed a comprehensive review of coaching and its impacts***). Many states want to use coaching throughout their technical assistance and quality improvement initiatives but it is not sustainable.

In order to deal with these problems of efficiency, a new technology called "Pinging" has been devised where training/professional development segments can be sent directly to a cell phone/tablet/computer based upon learning algorithms and where no face-to-face interaction is necessary. Everything occurs online with "pings" tied to an assessment of knowledge and/or behaviors that may be lacking which are then reinforced to become more positive. This is a new approach to coaching which is being monitored and evaluated as part of an NIH R01 grant (***iLookOut Child Abuse Prevention Training Program, B. Levi, PI***) to determine its efficacy, effectiveness, and efficiency.

Going beyond the professional development field there are some direct applications to learning and instruction in general. For example, could pinging be used as a means to individualize instruction and learning to help solve McVicker-Hunt's "Problem of the Match" or to address Vygotsky's "Zone of Proximal Development" via a skilled tutor? Could pinging be used as an individualized text for a learner in which based upon an assessment, only content relevant to the learner's strengths and weaknesses are presented to the learner's electronic device (laptop computer, tablet, smartphone). Rather than having standardized textbooks that reach maybe

50% of the students, let's have individualized texts that reach 100%; but doing it electronically rather than hard copy. Suddenly this technology could be efficient enough to make this happen. Having individualized texts as hard copy is not cost efficient and could never be sustained, but doing it electronically could be a game changer. It is differential learning rather than one-size-fits-all learning.

Conceptually, think of a bulls-eye with learning opportunities and content spread out all over the bulls-eye but few in the center of the bulls-eye. Now enter ping-pong where the learning opportunities and content can be targeted to just hit the center of the bulls-eye. This way we can optimize learning opportunities making them relevant to the specific learner which might not be the same learning opportunities for another learner who has a different profile of learning needs.

So what does ping-pong look like?

Think of the last time you took an exam and did really well on certain aspects of the exam but bombed others. Generally the instructor reviews all the right answers so you get the feedback on what you did wrong but that's where it ends. With ping-pong, you get an additional learning opportunity to extend learning about what you did not really understand with additional positive reinforcement giving you opportunities to test your knowledge further.

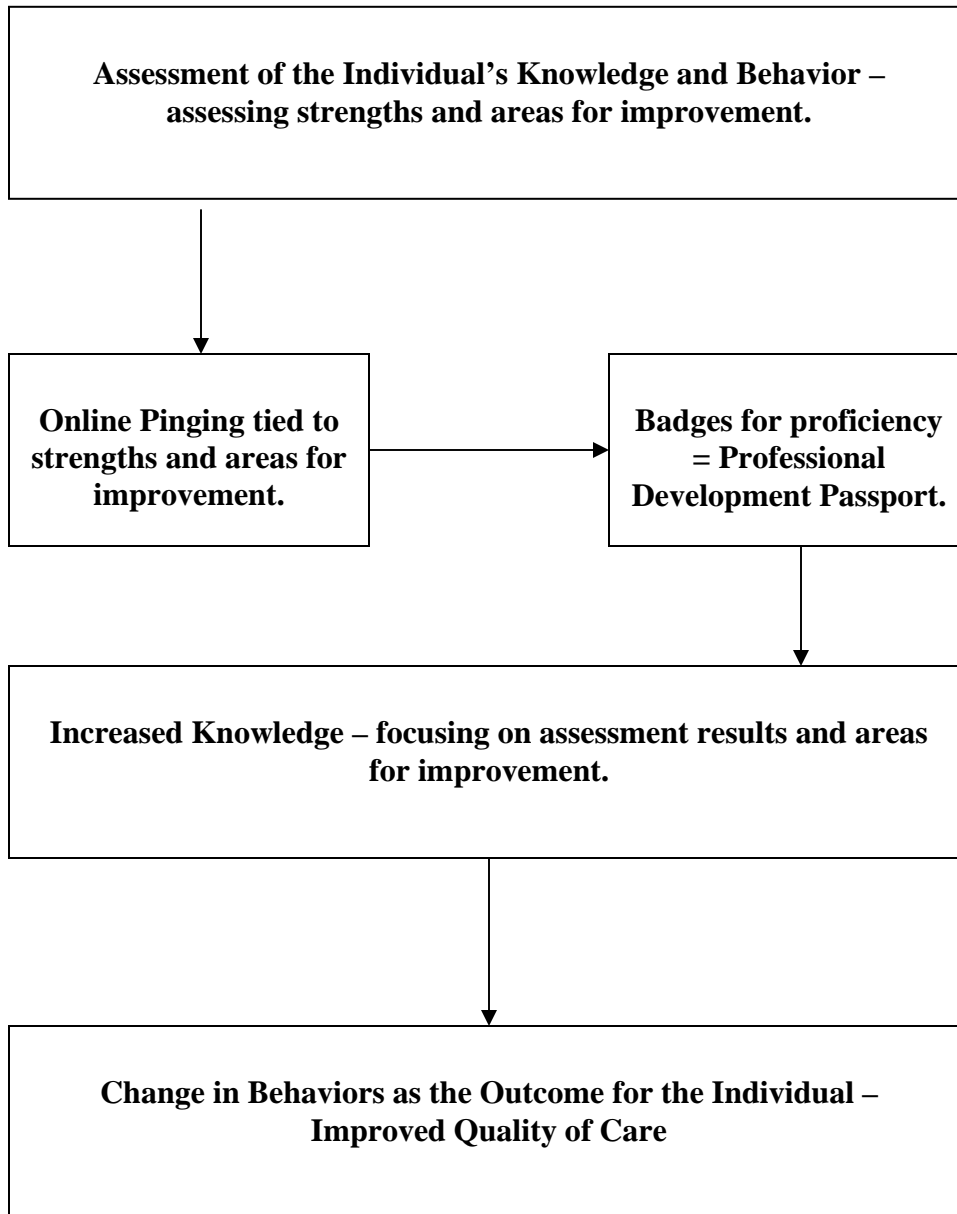
Algorithms are written that tie additional content to every exam question with additional supportive feedback which can be used to reinforce gaps in learning. These algorithms are activated based on the learner's test score. By doing this, we tie assessment to learning via ping-pong to give the individual learner the opportunity to enhance their learning beyond the assessment. In fact the assessment becomes the driver for additional learning via ping-pong rather than the assessment becoming the end goal. So rather than learn --> assessment we are changing the paradigm to learn --> assessment --> learn via ping-pong via multiple path ways. We are creating a learning - assessment - learning continuum. Here is the simple algorithm from the iLookOut program:

*Pre-Assessment --> iLookOut Learning Online Program --> Post-Assessment --> Pings sent*  
(A1, A2, A3, A4..) (B1, B2, B3, B4..) (C1, C2, C3, C4..) (D1, D2, D3..)

All this additional ping-pong occurs electronically sent to devices in a gamification format which becomes fun for the learner. It is cost efficient because the content is sent to a device without the need for a coach or instructor to follow up although that is always a possibility for a learner having a great deal of difficulty. An assessment can be done again after the ping-pong has occurred to determine the change in the learner's knowledge base. Other assessments could be used to see if behavior changes as well as knowledge changes have occurred depending on the content. For example, the NIH R01 grant we mentioned earlier is looking at just that, how knowledge changed about child abuse reporting, but also how it changed actual behaviors in reporting of child abuse, did it make for better reporting where false negatives and positives have decreased?

As we said at the beginning, this short paper or abstract is presented for its heuristic value to get us thinking about this new ping-pong technology as both a learning and coaching enhancement. The learning principles have been with us for some time, what is different now, is the available technology which could make a costly intervention more cost efficient. We have more questions about the technology than we have answers at this point. It has tremendous potential but we need to determine if it can live up to its billing as an effective and efficient enhancement.

**Problem Solving Coaching Equaling Online Pinging: Making Coaching Both Effective and Efficient (Fiene & Levi, 2017)**



Pinging Grant		Year 1 8/1/16–7/31/18	Year 2 8/1/17–7/31/18	Year 3 8/1/18–7/31/19	Year 4 8/1/19–7/31/20	Year 5 8/1/20–7/31/21
iLookOut	COMPLETE: Research study infrastructure		Sequential pings 1-52	Sequential pings 53-104	Sequential pings 105-156	Sequential pings (1-54) + Scenarios + Feedback (if optimal) + Leader Board (if optimal)
Standard	Content for iLookOut  Instrument validation	∅	Pings + Scenarios +/- Feedback Measure impact on: Judgment & Motivation	Pings + Scenarios + Feedback (Cross-over design) Measure impact on: Judgment & Motivation		
Control	Filming/editing  Recruitment materials	∅	∅	∅	Pings +/- Leader Board (No vs. Yes Person vs. Group) Measure impact on learner achievement	
State-wide	Pinging & badging plan  IRB approvals	∅	∅	∅		



DR RICHARD FIENE, RESEARCH PSYCHOLOGIST  
RESEARCH INSTITUTE FOR KEY INDICATORS, LLC (RIKI)  
1871 GRAMERCY PLACE  
HUMMELSTOWN, PENNSYLVANIA 17036  
MOBILE 717-598-8908  
FIENE@RIKINSTITUTE.COM  
RIKINSTITUTE.COM

PA ENTITY #: 4193691 D-U-N-S #: 011323030

## Better Kid Care Coaching + Online Modules ITERS Statistical Design

Richard Fiene

September 2017

The purpose of this technical research note is to outline the statistical design for evaluating the effectiveness of a coaching intervention and determining the specific relationships between key module content and ITERS indicators. The statistical design has two components: 1) t-tests to determine equivalency (pre-test) of two groups (Coaching and Comparison) and their subsequent sub-scale scores on a post-test after the coaching intervention has been administered. The comparison group will only receive the normal online modules that are readily available to all child care providers.

	<u>Pre-Test</u>	<u>Post-Test</u>
Coaching	C1	C2
Comparison	C3	C4

C1 + C3 should show non-significant differences on the ITERS scores. C1 -> C2 should show significant differences on the ITERS based on the coaching intervention. C3 -> C4 should show some significant differences on the ITERS but not as much as C1 -> C2. And lastly, C2 <-> C4 should show significant differences on the ITERS with C2 being significantly higher.

The second component of the statistical design is as follows: 2) correlations will be conducted between the specific online modules and the ITERS indicators (n = 420). Patterns or paths in the data will be determined to ascertain any relationships between how well classrooms did on the ITERS and what specific online module course content was taken.

	<u>ITERS Indicators</u>						
	<u>1.1.1</u>	<u>1.1.2</u>	<u>1.1.3</u>	<u>1.1.4</u>	<u>1.1.5</u>	<u>1.1.6</u>	<u>1.1.n.....</u>
Modules							
1	sign	ns	ns				
2				sign	ns		
3	sign						
4	ns	ns	ns	ns	ns		

By using the above statistical design, one can determine the effectiveness of the coaching intervention and specifically what modules were most effective.

# **FUTURE ANALYSES AND RESEARCH RELATED TO DIFFERENTIAL MONITORING, KEY INDICATORS, AND RISK ASSESSMENT METHODOLOGIES UTILIZING PREDICTIVE ANALYTICS**

**Richard Fiene**

**January 2015**

This short paper addresses what I see as the key future analyses and research related to differential monitoring, key indicators, and risk assessment methodologies. Most of these analyses can most likely be performed via predictive analytics.

## **Research Questions:**

1...There is the need to address the point system within the Differential Monitoring Scoring Protocol (DMSP©) by looking at the probability that the various key elements will occur based upon the research literature. For example, PC x PQ is .5 based upon NQI data because 50% of the states have QRIS systems. This is how all the algorithms would play out if a probability assessment is used rather than the scoring protocol I developed. The scoring protocol mirrors the probability figures as follows:

$$PC + PQ = .50P/4PTS$$

$$KI + RA \rightarrow DM = .50P/4PTS$$

$$PC + KI \rightarrow DM = .25P/2PTS$$

$$PC + RA \rightarrow DM = .25P/2PTS$$

2...There is the need to show how KI and RA are integrated mathematically or via an algorithm.

3...With the effectiveness and efficiency relationship curves (see my DMLMA Powerpoint slides). The effectiveness and efficiency lines are curvilinear rather than linear and cross each other at a substantial compliance level rather than earlier which is more typical with linear data.

4...HSKI as the best case example which incorporates all components. Full data sets, report, training slides, validation data, promotional slides, web site, most details and national DB. This needs to be documented fully and written up as a case study.

5...Run phi correlation against Logit regression, compare results.



6...2 x 2 phi to a 2 x 3 chi square. High/Low frequency matrix to Full/Substantial/Low frequency matrix.

7...ECPQIM/PAM/Measures = DM/Clustering/DMSP//KI/Classification/Matrix//RA/clustering/Likert.

There needs to be a paper written on the relationship between ECPQIM, predictive analytics modeling (PAM), and the actual measures used for each ECPQIM Key Element. I started this paper but it needs to be fully developed (see DATA File Folder).

8...Try different cut offs and see how results are impacted. I started to do this with the GA data base.

The more the indicators, the higher the correlation between IC and CI. KI8 --> KI15. The question becomes what is the best level? KI10, KI9, KI13??? This analysis ties back to the efficiency and effectiveness relationship because as one increases the number of indicators, the effectiveness increases but the efficiency of the model drops off. The opposite is also true.

9...Use HS/KS/IL/GA data bases to run the various analyses. These data bases are available for doing all these analyses.

10...DM = YES OR NO, BASED UPON COMPLIANCE HISTORY; H = YES (100-98); L = NO (97-); YES = KI AND/OR RA (ABBREVIATED INSPECTION); NO = CI (FULL INSPECTION); CLUSTERING OR CLASSIFICATION. These are the various key elements of ECPQIM and the types of analyses within predictive analytics modeling (clustering or classification analysis).

11...DMSP – 0-10; CLUSTERING (0,2,4,6,8,10). DMSP – Differential Monitoring Scoring Protocol is an example of clustering analysis via predictive analytics modeling.

12...KI -- .25+; CLASSIFICATION; either it is included or not. KI – Key Indicators is an example of classification analysis via predictive analytics modeling.

13...RA – 9 OUT OF 10 (9+); HIGH RISK; CLASSIFICATION; either it is included or not. RA – Risk Assessment is an example of classification analysis via predictive analytics modeling.