
Your EARLY CHILDHOOD COMMUNITY
Exchange™
SUMMER 2025
VOLUME 47, ISSUE 2, NO. 278
$15.00
ON OUR COVER
JCC Z Frank Apachi
Early Childhood Center
Northbrook, Illinois
FOCUS ON
Connections Across
Generations

Potential Solution to the Child Care Trilemma Revisited

Finding the "Right Rules"—The Holy Grail of Early Care and Education

by Richard Fiene

Rules and regulations: Can't live with them, Can't live without them.

How often have you heard this statement? I have heard it a great deal in an early care and education career that has seen six decades of discussion about what is the right mix of rules and regulations, the basic protections for children while in out of home child care. Recently, in the early care and education field, there has been a great deal of discussion about deregulation of early care and education standards/rules/regulations in order to have increased access to child care (National Association for the Education of Young Children: NAEYC, 2024). This discussion or controversy has been going on for a long time, it is nothing new. I remember back in the early 1970's when I was directing the Mary Elizabeth Keister Infant Toddler Demonstration Center at the University of North Carolina at Greensboro and there were discussions about the revision to the Federal Interagency Day Care Requirements (FIDCR). What was the right mix, the balance of protections and quality enhancements for young children in child care that the Federal Department of Health, Education and Welfare wanted to promulgate nationally.

Rick Fiene
Exchange Leadership Initiative
exchangepress.com/leadership

Rick Fiene has been working in prevention research for more than 50 years, focusing on child care policy, child care quality and human service regulatory administration and science. He initially started in prevention research because he wanted to have a positive impact on children's lives. He believed that researchers were not paying enough attention to the licensing of child care programs and decided to base his career on improving this area of research.

Fiene is regarded as a leading international researcher/scholar on human services licensing measurement and differential monitoring systems. His theory of regulatory compliance has altered human services regulatory science and licensing measurement dramatically in thinking about how best to monitor and assess licensing rules and regulations through targeted and abbreviated inspection methodologies: differential monitoring, risk assessment, and key indicators. He received the National Association for Regulatory Administration President's Award in 2015 and the Pennsylvania Association for the Education of Young Children's Distinguished Career Voice for Children Award in 2020.

But I think there is a better way to deal with this discussion which is driven by regulatory science and the empirical evidence that has emerged over the past 50 years. Let's take this discussion out of the political domain and place it where it needs to be, firmly within the newly emerging regulatory science field and focus on regulatory compliance. There is a theory of regulatory compliance (Fiene, 2019) getting kicked around a good deal in the human services regulatory science field that has upended the way we make licensing decisions. The theory has been empirically proven in several studies throughout the U.S. and Canada (Fiene, 2025). The theory simply states that substantial regulatory compliance with child care rules and regulations may be equivalent to full (100 percent) regulatory compliance with all child care rules and regulations. From a public policy and licensing decision making point of view, it changes program monitoring from a uniform one-size-fits-all approach to a more targeted and focused differential monitoring approach that looks at risk assessment and prediction of overall compliance with rules and regulations (Fiene, 2025).

So that is the theory but where do we start at a practitioner level? If we start at the baseline of early care and education

quality, then licensing and *Caring for Our Children (CFOC): The National Health and Safety Performance Standards, 4th Edition* (AAP, APHA, NRCHSCC, 2019), published by the American Academy of Pediatrics (AAP), the American Public Health Association (APHA), and the National Resource Center for Health and Safety in Child Care (NRCHSCC) is a good place to start because the CFOC is considered the default set of health and safety standards in the early care and education field. The standards were first published in the early 1990's and have been refined through several revisions and editions over the past several decades in response to the everchanging early care and education research literature related to health, safety and program quality. For over 30 years, the standards have been the reference for state child care licensing agencies as they think about promulgating new or revised rules/regulations/standards in their respective states. It is based upon the latest science in developmental psychology, pediatrics, and public health fields related to early care and education settings. The standards have been peer reviewed by expert technical panels representing all of the above areas of developmental psychology, pediatrics, public health, environmental health, etc. But it is a daunting document, over 700 standards are within this reference manual for the early childhood field.

Advocates point to *Caring for Our Children* (AAP, APHA, NRCHSCC, 2019) as the go-to-document because it provides a solid floor to quality while building on this base to demonstrate best practices. Others, mostly in the political arena, point to it as an example of over-regulation, too many rules to follow. But let's not forget what *Caring for Our Children* (AAP, APHA, NRCHSCC, 2019) is all about, protecting our children while in out of home care. Access to child care is important for many families, as is access to quality child care, as is access to safe and quality child care. Navigating these all is a delicate and challenging balance.

So, what is a potential solution to the child care trilemma? Let's look at regulatory science for potential guidance. As I said earlier, regulatory science is an emerging field, it is not well developed as the other physical and social sciences but it is making tremendous strides in the past 20-30 years. There are two parallel tracks, one dominated by the pharmaceutical industry and the other in the human services, in particular, in early care and education. In the pharmaceutical arena there is more concern about clinical trials and the efficacy of drugs and protection from side effects for individuals; in the human services arena there is more concern about protections from harm related to caregiving. And this is where regulatory science came into play with a new methodology in the human services that was emerging around risk assessment and key indicator rules/regulations to make monitoring more effective and efficient by focusing on risk to children and prediction of overall regulatory compliance (Fiene, 2019, 2025).

Initially there was more focus on the risk assessment methodology to determine if certain *Caring for Our Children* (AAP, APHA, NRCHSCC, 2019) standards placed children at increased risk of morbidity and mortality if regulatory non-compliance occurred. The resulting document, *Stepping Stones to Caring for Our Children* (NRCHSCC, 2019), came about based upon this risk assessment rule methodology. It took the over 700 standards to distill it down to approximately 120 standards. It became a much more manageable document that state licensing agencies could use as a guide in revising and promulgating rules and regulations.

Later in the development and evolution of *Stepping Stones to Caring for Our Children* (NRCHSCC, 2019), again borrowing from the regulatory science field, the key indicator rule methodology was utilized to determine if there were a smaller set of standards that had more of a predictive value in protecting children when it came to regulatory compliance in an overall sense. This resulted in *Caring for Our Children Basics* (CFOCB) (ACF, 2015) (approximately 65 standards) which was originally proposed as a voluntary set of standards for all early care and education. I think it was a good idea back when it was first proposed in 2015 and I still think it is a good idea. To many, 65 standards may still sound like too many standards but these standards form the basis for the quality and safety arm when it comes to the child care trilemma, and indirectly impact accessibility and affordability. The more standards to meet, the greater the cost for programs which can make it more difficult for parents to access available care. As quality increases, so does cost while accessibility decreases based upon what parents can afford.

Let's begin here in attempting to address a revised solution to the child care trilemma. In this discussion about where the child care field is headed and the most recent call for deregulation (Hechinger Report, 2024)(NAEYC, 2024), let's pivot and think about using *Caring for Our Children Basics* (ACYF, 2015) as our point of discussion rather than arbitrarily removing rules with this deregulation mind set because it is politically expedient. Let's be driven by the empirical evidence and the science which *Caring for Our Children Basics* (ACYF, 2015) is derived from solid regulatory compliance methodologies of risk assessment and key indicator rule/regulatory/standard identification. See how your state's child care rules size up with *Caring for Our Children Basics* (ACYF, 2015) in making sure that at the very least all these standards are in place. Templates from regulatory science have been developed

to do this comparison (Fiene, 2025). As a very important footnote regarding these standards, they were developed by a cross-representation of medical experts, early care and education experts, child developmental experts, public health and environmental experts. So all disciplines having an impact on child care services were well represented and consulted in the development of the standards.

Then once this is done in the aggregate, begin to look at the individual standards within *Caring for Our Children Basics* (ACYF, 2015). Let's be honest, probably the most discussed standard is staff-child ratios and group sizes. It has the greatest impact on cost (staff), numbers (children), and quality. This has been clearly demonstrated in the research literature for over 50 years. Nothing has changed, it was the focal point back in the 1970's (Abt, 1979) and it is today (Fiene & Stevens, 2021). But let's think outside the regulatory compliance box for a minute and maybe we do not look at staff-child ratios in isolation but cross it with another standard/rule such as the qualifications of staff and suggest an alternate rule where staff-child ratio can be increased slightly but only with the most highly qualified staff?! Like I said, let's think outside the regulatory compliance box. And while we are there, the fee that is attained by the program with the additional child should go to the more qualified staff as an add on to their salary. Yes, they have an additional child but they also have the revenue generated as a salary increase with the addition. This above approach I suggested in a Child Care Information Exchange article back in 1997 in how this approach could be utilized effectively as a potential solution to the child care trilemma (Fiene, 1997).

As with staff-child ratio and group size, we perform the same type of critical analysis utilizing the empirical regulatory compliance data available to make changes in the existing set of rules. As has been pointed out in the regulatory science research literature, regulatory compliance with rules is a measurement issue, so it should be solved in a corresponding way, use the data, do not ignore the empirical evidence and leave it up to the whims of the political process to determine what stays and what gets pitched. For the interested reader, there are several studies completed by the National Association for Regulatory Administration (NARA) which can guide one in determining how best to use data to make these decisions. These can be found at naralicensing.org/key-indicators

The point of this research abstract position paper is for us to take a step back and avoid a knee-jerk reaction to dealing with the child care crisis and that the only solution is to increase availability and affordability at the expense of health, safety and quality via deregulation (NAEYC, 2024)(Hechinger Report, 2024). We now have an emerging regulatory science (Fiene, 2025) to guide us and I hope we use it for making educated and informed choices as we move forward in attempting to solve the continuing child care trilemma.

References

Abt Associates. (1978). *Day Care Centers in the U.S.A National Profile 1976-1977*, Boston, Massachusetts.

American Academy of Pediatrics and the American Public Health Association and the National Resource Center for Health and Safety in Child Care: AAP, APHA, NRCHSCC. (2019). *Caring for Our Children: The National Health and Safety Performance Standards*. Itasca, Illinois.

Administration for Children and Families: ACF. (2015). *Caring for our Children Basics*. Washington, DC.

Fiene, R. (2025). Finding the Rules that Work: An emerging paradigm promised to close the gap between regulatory compliance scores and quality of childcare services. *American Scientist, 113* (January-February 2025).

Fiene, R. (2024). *Regulatory Compliance and Monitoring Systems*. Research Institute for Key Indicators and National Association for Regulatory Administration.

Fiene, R. (2019). A Treatise on the Theory of Regulatory Compliance. *Journal of Regulatory Science, 7*, 1-3.

Fiene, R. (1997). Searching for a Solution to the Child Care Trilemma. *Child Care Information Exchange*, September, 1997, 57-59, Redmond, Washington.

Fiene, R. & Stevens, S. (2021). *Contact Hours as a New Metric Replacing Group Size and Staff-Child Ratios as well as a New Metric for COVID19 Thresholds*. Research Institute for Key Indicators/Penn State.

Hechinger Report. (2024). *Investigating How Kids are Harmed by Lower Child Care Standards*. New York, New York.

National Resource Center for Health and Safety in Child Care: NRCHSCC. (2019). *Stepping Stones to Caring for our Children*. Boulder, Colorado.

National Association for the Education of Young Children: NAEYC. (2024). *The Costs of Deregulating Child Care*. Washington, DC.

knowledge

Study Protocol

The Uncertainty–Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies

Richard Fiene

MDPI

https://doi.org/10.3390/knowledge5020008

knowledge

MDPI

Study Protocol

The Uncertainty–Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies

Richard Fiene

Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, PA 16802, USA; rfiene@rikinstitute.com

Abstract: This research article proposes the use of an uncertainty–certainty matrix (UCM) for licensing decision making in the human services, which is the decision to issue a license to operate. It is a proposed study protocol and conceptual framework; it is not an empirical study. It shows how the matrix can be used in rule decision making and how it clearly shows when decision making has gone awry when bias is introduced into the decision making. It is also proposed to be used to make decisions in differential monitoring and in validation and reliability studies. This proposal presents a potential blueprint on how the UCM can be used within human services licensing as a decision-making tool.

Keywords: decision making; uncertainty–certainty matrix; regulatory compliance; licensing; reliability and validation studies

Academic Editor: Cengız Kahraman

Received: 1 January 2025
Revised: 25 April 2025
Accepted: 27 April 2025
Published: 28 April 2025

Citation: Fiene, R. The Uncertainty–Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies. *Knowledge* **2025**, *5*, 8. https://doi.org/10.3390/knowledge5020008

1. Introduction

This research proposal takes the Contingency Table, which is a well-known metric in the statistical decision-making research literature [1], and refocuses it on regulatory science within the context of the definition of regulatory compliance and licensing measurement. It also deals with the policy implications of this particular metric. In this study protocol, it is proposed that the Uncertainty–Certainty Matrix (UCM) is a fundamental building block to licensing decision making from a measurement perspective. The Contingency Table, as demonstrated by a 2 × 2 matrix, is utilized in regulatory compliance and is the center piece for determining licensing key indicator rules [2], but it is also a core conceptual framework in licensing measurement and ultimately in program monitoring and reviews [3].

The reason for selecting this matrix is the nature of licensing data: it is binary or nominal in measurement. Either a rule/regulation is in compliance or out of compliance. Presently, most jurisdictions deal with regulatory compliance measurement in this nominal level or binary level. There is to be no gray area; this is a clear distinction in making a licensing decision about regulatory compliance. The UCM also takes the concept of Inter-Rater Reliability (IRR) a step further in introducing an uncertainty dimension that is very important in licensing decision making which is not as critical when calculating IRR. Inter-Rater Reliability is a real concern in the human services licensing field in that in many cases it is difficult for jurisdictions to maintain a high degree of consistency when comparing individual licensing inspectors to each other. Part of the problem is a fundamental measurement issue; it is hoped that the addition of the UCM will help to mitigate this problem [4]. Licensing measurement is dominated by nominal measurement: either a rule is in compliance or it is out of compliance. A proposal has been suggested in which an ordinal scale based upon licensing rule violations would be utilized called the

Regulatory Compliance Scale (RCS) [3]. This new RCS scale shows promise, but it needs additional validation studies in order to be used on a regular basis for making human services licensing decisions (2a).

The reason for even suggesting this matrix is the high level of dissatisfaction with the levels of reliability in the results of program monitoring reviews as suggested in the previous paragraph. If the dissatisfaction was not so pronounced, it would not be an issue, but with it being so high, the field of licensing needs to take a proactive role in determining the best possible way to deal with increasing inter-rater reliability among licensing inspectors. Hopefully, this organizational schema via the UCM Matrix will help to think through this process related to licensing measurement and monitoring systems. The author has been conducting regulatory compliance studies for the past 50 years and has determined that the validity and reliability of these studies needs a more robust model for making licensing decisions via more accurate measurements of regulatory compliance. This led to the creation and proposing of the UCM Metric [5–7].

Over the past 50 years, it has been well documented by the National Association for Regulatory Administration (NARA) how the licensing field has changed in moving from a one-size-fits-all licensing and monitoring approach to one of differential or targeted licensing and monitoring (https://www.naralicensing.org/key-indicators, accessed on 24 April 2025). NARA has led this transition in the human services licensing and regulatory administration field, which has produced a much more productive, effective, and efficient licensing inspection system. The UCM and RCS are the latest pieces in the puzzle to accomplishing this new licensing decision-making framework.

The key pieces to the UCM are the following: the decision (D) regarding regulatory compliance and actual state (S) of regulatory compliance. Regulatory Compliance of individual Rules: Plus (+) = In-compliance, or Minus (−) = Out of compliance. As such, the matrix can be built as follows (Table 1):

Table 1. Uncertainty–Certainty Matrix (UCM) Logic Model.

UCM Matrix Logic		**Decision (D) Regarding**	**Regulatory Compliance**
		(+) In Compliance	(−) Not In Compliance
Actual State (S) of Compliance	(+) In Compliance	Agreement	Disagreement
	(−) Not In Compliance	Disagreement	Agreement

The above UCM matrix demonstrates when agreement and disagreement occur, which establishes a level of certainty (Agreement Cells) or uncertainty (Disagreement Cells). In a perfect world, there would only be agreements and no disagreements between the decisions made about regulatory compliance and the actual state of regulatory compliance. However, from experience, this is not the case. This is based up reliability testing carried out in the human services licensing research field in which a decision is made regarding regulatory compliance with a specific rule or regulation, and then that is verified by a second licensing inspector observer who generally is considered the measurement standard.

Disagreements raise concerns in general, but the disagreements are of two types: false positives and false negatives. A false positive is when a decision is made that a rule/regulation is out of compliance when it is in compliance. This is not a good thing, but its twin disagreement is worse. With false negatives, it is decided that a rule/regulation is in compliance when it is out of compliance. False negatives need to be avoided because they place clients at extreme risk more so than a false positive. False positives should also be avoided, but it is more important to deal with the false negatives first before addressing the false positives.

2. Uncertainty–Certainty Matrix for Validation and Reliability Studies

This part of the research proposal is to explore the possibility of utilizing the Uncertainty–Certainty Matrix (UCM) as depicted in Table 1 in validation and reliability studies in licensing decision making. The UCM has been proposed for use in licensing decision making, but this would be an extension of this thinking to studies that involve validating licensing decisions, such as when key indicators/predictor rules are used in comparison with comprehensive reviews of rules [5] and in reliability studies to determine individual inspector bias in regulatory compliance [8,9].

The basic premise of the UCM is that individual decision making matches reality. When it comes to regulatory compliance decision making. a 2 × 2 matrix can be drawn with the possible outcomes as indicated in the following table (Table 2), which is based upon the logic of Table 1.

Table 2. Uncertainty–Certainty Matrix (UCM) Logic Model applied to Validation Studies.

UCM Matrix Logic	For Validation Studies	Decision Regarding	Regulatory Compliance
		(+) In Compliance	(−) Not In Compliance
Actual State of Compliance	(+) In Compliance	Agreement (++)	Disagreement (+−)
	(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

In using this table, the hope is that the decision regarding regulatory compliance matches the actual state of compliance where the coefficient is as close to +1.00 as possible; in other words, perfect agreement. The agreement cells are heavily weighted (++) and (−−). We do not want to see all the cells, both agreement and disagreement cells, equally weighted (++), (+−), (−+), (−−). That would indicate a random response rate and a coefficient close to 0.00.

However, there is another possibility which involves bias on the part of the licensing inspector in which they have certain biases or tendencies when it comes to making regulatory compliance decisions about individual rules. Consequently, it is possible that decisions made regarding regulatory compliance could be either overall (+) positive In-Compliance or (−) negative Not-In-Compliance when in reality, the actual state of compliance is more random.

The UCM can be used for both reliability and validity testing as suggested in the above table (Table 2). For validity, false positives (+−) and negatives (−+) should either be eliminated or reduced as well as possible, and the remaining results should show the typical diagonal pattern as indicated by the agreement cells.

For reliability, the same pattern should be observed as in the validity testing above, but there is an additional test in which bias is tested for. Bias is ascertained if the patterns in the results indicate a horizontal or vertical pattern in the data with little or no diagonal indication. Bias can be found at the individual inspector level, as well as at the standard level or the actual state of compliance. This could provide a helpful visual for licensing administrators regarding how decisions are being made about program regulatory compliance in the field.

In both reliability and validity testing, random results in which each of the cells are equally filled are not desirable either. Obviously, additional training involving licensing inspectors would need to occur in order to make the data collection efforts both reliable and valid. Monitoring of regulatory compliance history data would need to occur on an ongoing basis to make sure that biases did not return or if new biases developed within the regulatory compliance system.

The following Tables 3–8 depict the above relationships with results highlighted in red:

Table 3. Valid and Reliable Results.

Valid and Reliable Results	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Table 4. Random Results.

Random Results	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Table 5. Positive Bias Results Individual Assessor.

Positive Bias Results Individual	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Table 6. Negative Bias Results Individual Assessor.

Negative Bias Results Individual	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Table 7. Positive Bias Results Standard.

Positive Bias Results Standard	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Table 8. Negative Bias Results Standard.

Negative Bias Results Standard	**(+) In Compliance**	**(−) Not In Compliance**
(+) In Compliance	Agreement (++)	Disagreement (+−)
(−) Not In Compliance	Disagreement (−+)	Agreement (−−)

Tables 3–8 demonstrate the different results based upon individual response rates when making regulatory compliance decisions about rules. Table 3 is what needs to be attained and Tables 4–8 need to be avoided. Only in Table 3 are false negatives and positives eliminated or avoided. In Tables 4–8, false negatives and/or false positives are introduced, which is not desirable when making validity or reliability decisions.

Table 4 results clearly indicate that a great deal of randomness has been introduced in the regulatory compliance decision making in which the individual licensing inspector decisions do not match reality. Tables 5 and 6 demonstrate bias in the decision-making process either positively (inspector always indicates in compliance) or negatively (inspector always indicates out of compliance). It is also possible that the standard being used has bias built into it; this is less likely but is still a possibility. The results in Tables 7 and 8 demonstrate where this could happen.

All these scenarios need to be avoided and should be monitored by agency staff to determine if there are patterns in how facilities are being monitored.

3. Uncertainty–Certainty Matrix for Differential Monitoring Studies

The purpose of this part of this research proposal is to explore the possibility of utilizing the Uncertainty–Certainty Matrix (UCM) not only in validation and reliability studies in licensing decision making, but also with differential monitoring studies. The UCM has been proposed for use in licensing decision making, but this would be an extension of this thinking to studies that involve validating licensing decisions, such as when key indicators are used in comparison with comprehensive reviews of rules and in the development of risk rules as part of the risk assessment methodology [4]. This new Differential Monitoring 2 × 2 Matrix can also be used to depict the relationship between full and substantial regulatory compliance and the nature of rulemaking.

The basic premise of the DMM: Differential Monitoring Matrix is similar to the original thinking with the UCM Matrix Logic as depicted in Table 1, but there are some changes in the formatting of the various cells in the matrix (see Table 9). When it comes to regulatory compliance decision making, a 2 × 2 matrix can be drawn with the possible outcomes as is indicated in Table 9 where each individual rule is either in (+) or out (−) of compliance. Additionally, there is the introduction of a high regulatory compliant group (+) and a low regulatory compliant group (−), which is different from the original UCM.

Table 9. DMM—Differential Monitoring Matrix.

DMM Matrix	High Group (+)	Low Group (−)
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

By utilizing the format of Table 9, several key components of differential monitoring can be highlighted, such as key indicators and risk assessment rules, as well as the relationship between full and substantial regulatory compliance.

Regulatory compliance is grouped into a high group (+); generally, this means that there is either full or substantial regulatory compliance with all rules. The low group (−) usually has 10 or more regulatory compliance violations [4]. Individual rules being in (+) or out (−) of regulatory compliance is self-explanatory.

Tables 10–16 below demonstrate the following relationships:

Table 10. Key Indicators/Predictor Rules.

Key Indicators	High Group (+)	Low Group (−)
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 11. Risk Rules/Place Clients at Increased Risk.

Risk Rules	High Group (+)	Low Group (−)
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 10 depicts the key indicator relationship between individual rules and the high/low groups as indicated in red. In this table, the individual rule is in compliance with

the high group and is out of compliance with the low group. This result occurs on a very general basis and should have a 0.50 coefficient or higher with a p value of less than 0.0001.

Table 12. Full Compliance with All Rules.

Full Compliance	**High Group (+)**	**Low Group (−)**
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance		(−−)

Table 13. Substantial Compliance with All Rules.

Substantial Compliance	**High Group (+)**	**Low Group (−)**
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 14. Very Difficult Rules.

Very Difficult Rule	**High Group (+)**	**Low Group (−)**
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 15. Poor Performing Programs.

Poor Performing Programs	**High Group (+)**	**Low Group (−)**
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 16. Terrible Rule.

Terrible Rule	**High Group (+)**	**Low Group (−)**
(+) Rule is In Compliance	(++)	(+−)
(−) Rule is Not In Compliance	(−+)	(−−)

Table 11 depicts what most rules look like in the 2 × 2 DMM. Most rules are always in full compliance since they are standards for basic health and safety for individuals. This is especially the case with rules that have been weighted as high-risk rules. Generally, one never sees non-compliance with these rules. There will be a substantial number of false positives (+−) found with high-risk rules, but that is a good thing.

Table 12 depicts what happens when full compliance is used as the only criterion for the high group. Notice that the cell right below (++) is eliminated (−+). This is highly recommended since it eliminates false negatives (−+) from occurring in the high group. As is seen in Table 12, when substantial compliance is used as part of the high group sorting, false negatives are re-introduced. If possible, this should be avoided; however, in some cases, because of the regulatory compliance data distribution, this is not always possible where not enough full compliant programs are present.

Table 13 depicts what occurs when substantial compliance is used as part of determining the high group. False negatives can be reintroduced into the matrix which needs to be either eliminated or reduced as best as possible. If substantial compliance needs to be used in determining the high group, then there is a mathematical adjustment that can be made, which will impact the equation and essentially eliminate false negatives mathematically.

Table 14 depicts what happens if the individual rule is particularly difficult to comply with. Both the high performers as well as the low performers are out of compliance with the rule.

Table 15 depicts a situation where the programs are predominantly in a low group with few at full or substantial regulatory compliance, which is indicative of poor performing programs. Very honestly, this is generally not seen in the research literature, but it is a possibility and one to be in tune with.

Table 16 depicts a terrible individual rule which predicts just the opposite of what we are trying to do with programs. Obviously, this rule would need to be rewritten so that it fits with the essence of regulatory compliance in helping to protect individuals.

The following Tables 10–16 depict the above relationships with results highlighted in red.

Tables 10–16 demonstrate the different results based on the relationship between individual regulatory compliance and if a program is either a high performer or a low performer. These tables are provided as guidance for understanding the essence of differential monitoring and regulatory compliance, which has various nuances when it comes to data distribution. This research proposal for a UCM hopefully can be used as a guide in determining from a data utilization point of view how to make important regulatory compliance policy decisions, such as which rules are excellent key indicator rules, which are performing as high risk rules, the importance of full compliance, what to do when substantial compliance needs to be employed, are there difficult rules to comply with, how well are programs performing, and do we have less than optimal rules that are in need of revision.

4. Conclusions

The Uncertainty–Certainty Matrix (UCM) should provide a useful tool for assessing the effectiveness of licensing decision making in the human services via validation and reliability studies within differential monitoring systems by visually inspecting cell proportions to determine if the appropriate results are depicted in the above matrices.

It is hoped that licensing researchers and regulatory scientists will experiment with it and test it out in different arenas beyond early care and education programs. It appears to have broad applicability across regulatory disciplines. The matrix has helped to identify the need to address false positives and negatives in the human services licensing decision-making process which undermines the effort of protecting clients.

The UCM also appears to provide a framework to identify reliability issues across licensing inspectors carrying out evaluations of individual programs. This issue of reliability is a big issue in the human services licensing field where there is a great deal of inconsistency when it comes to measuring regulatory compliance [10–12]. The UCM could be applied to existing regulatory compliance history data to determine if bias is present or not. It provides a clear visual demonstration of when regulatory compliance history data have gone awry and are not performing as they should. This can be a useful tool for licensing administrators in making changes to their overall licensing system, as well as for which individual rules/regulations/standards are most effective in protecting clients or might need revision.

The major limitation of the UCM is that as of this writing, it has not been empirically tested to see if this conceptual framework is really helpful to licensing policymakers and researchers. The UCM is a theoretical model at this point that needs to be verified. At the same time, it holds promise for the human services licensing field because the field as it relates to regulatory science has a measurement problem when it comes to reliability and

validity. Without a solid measurement structure, it is the old adage of "Garbage In, Garbage Out". Hopefully, the UCM is a first step to rectifying this issue.

Clearly, for future research, there needs to be additional expansion beyond the child-care and early education field to all of human services and then beyond this scope to other regulatory areas to determine if a UCM approach is relevant. It is obvious that in clinical studies within the medical field that the UCM would be very appropriate in order to avoid false negatives where a drug's side effects would be more detrimental than the potential benefits from taking the particular drug. We need additional real-life examples where the UCM model can be tested to see how useful it would be in other regulatory settings.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pearson, K. On the Theory of Contingency and Its Relation to Association and Normal Correlation. In *Draper's Company Research Memoirs, Biometric Series I*; Cambridge University Press: London, UK, 1904.
2. Fiene, R. A Treatise on Regulatory Compliance. *J. Regul. Sci.* **2019**, *7*, 1–3. [CrossRef]
3. Fiene, R. *The Development of a Regulatory Compliance Scale*; Research Institute for Key Indicators/Penn State University: University Park, PA, USA, 2025.
4. Miranda, B.; Ekyalongo, Y.; Franchett, A.; Maxwell, K. Monitoring Practices Used in Child Care and Early Education Licensing. In *OPRE Report #2022-137*; Office of Planning, Research, and Evaluation, Administration for Children and Families, Department of Health and Human Service: Washington, DC, USA, 2022.
5. Fiene, R. *Regulatory Compliance*; National Association for Regulatory Administration: Fredericksburg, VA, USA, 2024.
6. Trivedi, P.A. *Innovation in Monitoring in Early Care and Education: Options for States*; Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services: Washington, DC, USA, 2015.
7. Fiene, R. Finding the Rules that Work: An emerging paradigm promised to close the gap between regulatory compliance scores and quality of childcare services. *Am. Sci.* **2025**, *113*, 16–21.
8. Maxwell, K.; Miranda, B.; Harris, P.; Johnson, N.; Padilla, C.; Banghart, P.; Verhoye, A.; Early, D.M. Child care and early education licensing literature review. In *OPRE Report #2024-199*; Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services: Washington, DC, USA, 2024.
9. Zellman, G.; Fiene, R. Validation of quality rating and improvement systems for early care and education and school-age care. In *OPRE and Child Trends*; Office of Planning Research and Evaluation Administration for Children and Families, U.S. Department of Health and Human Services: Washington, DC, USA, 2012.
10. Lahti, M.; Elicker, J.; Zellman, G.; Fiene, R. Approaches to validating child care quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Early Child. Res. Q.* **2014**, *30*, 280–290. [CrossRef]
11. Fiene, R.; Kroh, K.E. *Human Services Licensing Measurement, Regulatory Compliance, and Program Monitoring Systems*; Research Institute for Key Indicators & the National Association for Regulatory Administration: Minneapolis, MN, USA, 2016.
12. Payne, A.L. *Strong Licensing: The Foundation for a Quality Early Care and Education System*; National Association for Regulatory Administration: Minneapolis, MN, USA, 2011.

A reprint from

American Scientist

the magazine of Sigma Xi, The Scientific Research Honor Society

Finding the Rules that Work

An emerging paradigm promises to close the gap between regulatory compliance scores and the quality of childcare services.

Richard Fiene

An old fable recounts how a father and son, taking a donkey to market to sell it, encounter a string of critical villagers who each inform the pair they're "doing it wrong." Their efforts to please each subsequent critic end, absurdly and tragically, with them carrying the beast of burden themselves, ultimately causing its death.

Like the advice of those villagers, regulations are proffered in the name of safety and good practice. And, like that father and son, programs that try to follow every single rule to the letter may soon find themselves too weighed down to achieve (or perhaps even recall) what they set out to do. As the saying goes, "When you're up to your behind in alligators, it's hard to remember that you set out to drain the swamp."

In my four decades as a regulatory scientist studying childcare, I've seen this pattern play out time and again: In the lead-up to evaluations, staff at perfectly compliant programs spend so much time dotting i's and crossing t's that they have little left over for working with classrooms or teachers, whereas staff at slightly less compliant facilities, though equally careful about observing rules, fuss less with paperwork and work more with teachers on improving skills and curriculum.

Needless to say, developmentally appropriate curricula change kids' lives; boasting a perfect record does not. This observation neither dismisses the 200 to 400 rules and regulations set by respective U.S. states nor undermines the importance of complying with them, either as individual rules or in the aggregate. And full compliance does improve safety. But, as data gathered by my research team repeatedly demonstrates, a vague, uncomfortable gap separates full, costly regulatory compliance from program quality.

It is never about more or fewer rules; it is about which rules are really productive and which are not.

Moreover, early care and education providers often voice concerns that licensing inspectors inconsistently administer and apply particular rules. At issue, then, are not regulations' overall value per se, but rather the value of individual rules relative to fanatical box-checking. Given their limited resources, how can the early care and education fields get the most bang for their buck?

Such a discussion is long overdue. The unequal worth of many general licensing and quality standards, including those driven by a regulatory political bent rather than empirical evidence, produce markedly uneven developmental outcomes for kids. Today, an outcomes-based scientific reference frame is already influencing the human services industry (childcare, child welfare, and child and adult residential services), particularly in the early care and education fields (childcare centers and family childcare homes for children between infancy and 12 years old). The point of my team's approach, which I call the *theory of regulatory compliance*, is not to ask whether we need more or fewer rules, or more thorough or less thorough compliance, but rather to evaluate which rules truly prove effective.

Modernizing Measurement

Regulatory scientists use tools, standards, and methodologies to assess the safety, efficacy, and quality of programs under government regulation. Ideally, they help regulatory agencies achieve the best possible public health and safety outcomes.

The regulatory science field has a lot of ground to make up. At about 30 years old, it lags its subject matter by a good century (Pennsylvania passed the first orphanage licensing law in the United States almost 140 years ago). Human services licensing grew slowly prior to the late 1960s to early 1970s, when American President Lyndon B. Johnson began the Great Society initiatives such as Head Start, which kicked off the rapid multipli-

QUICK TAKE

Contrary to historical assumptions, the quality of childcare programs does not increase linearly as their compliance with rules and regulations approaches 100 percent.

All-or-nothing, one-size-fits-all approaches to compliance and licensing generate skewed data, raise risks of false negatives and false positives, and burden staff with bureaucratic tasks.

Substantial regulatory compliance is an alternative approach that emphasizes compliance with the most productive rules, preserves safety, and allows staff to concentrate more on children.

DGLimages/Shutterstock

Staff of fully compliant childcare programs say they spend too much time box-checking and not enough working with teachers, whereas staff at slightly less compliant facilities, though equally scrupulous, bother less with form-filling and spend more time in the classroom. An outcomes-based substantial regulatory compliance approach lets licensors strike that balance.

cation of childcare programs. Those decades also saw human services, especially childcare, begin transforming from cottage industries, with program monitoring and measurement conducted qualitatively via case notes and anecdotal records, to more rigid systems that entailed oversight, case reviews, and state agency inspections. In the 1970s, these systems, which often varied from state to state, gave way to improvements brought by the Federal Interagency Day Care Requirements.

The watershed moment for regulatory science as it pertains to children's programs came in the 1980s. The previous decade's major childcare expansion in the United States had created a backlog of licensing assessments, caused unmanageable monitoring delays, and laid bare the logistical limits of case studies. These factors, combined with advances in computing, led states to introduce an empirical, quantitative, and instrument-based approach, complete with sophisticated software systems designed by state agencies and private vendors to track regulatory compliance and quality assessment data. Empirical evidence not only moved regulatory science from qualitative to quantitative analysis, it also revealed surprising patterns.

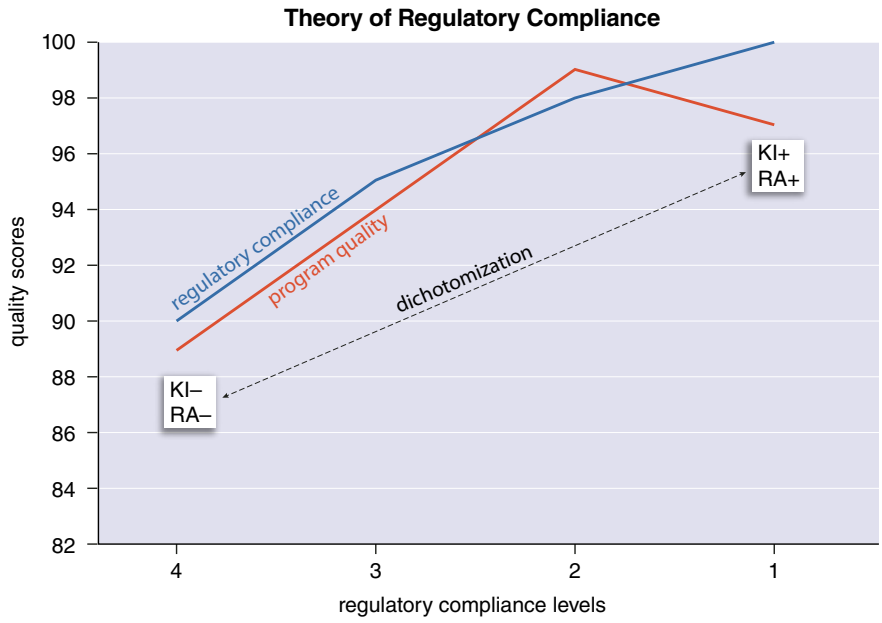
But first, some background: As the U.S. Department of Health, Education, and Welfare took over running the show for all U.S. early care and education programs in the 1970s, *uniform program monitoring* had become the rule. Uniform monitoring derived from the philosophical assumption that fuller regulatory compliance would produce, linearly, better quality across U.S. early care and education programs. As the former went up, so would the latter. From a public policy standpoint, this notion sounds aspirational, but sensible: Any licensing agency looks for service quality to increase as its rules, regulations, and standards are followed.

But as expert opinion and anecdotal evidence gave way to better-designed studies and empirical data, and as larger studies became possible thanks to data computerization by state licensing agencies, cracks appeared. When researchers compared violations found during licensing reviews and inspections to the quality of the violating programs, they found that a linear relationship did indeed exist between quality and compliance—but only as one moved from low compliance levels to *substantial regulatory compliance* (that is, 98–99 percent). Between that and 100 percent compliance, quality consistently plateaued and, as some 2010s replication studies suggested, even showed diminishing returns.

A New Paradigm

These results called into question the notion that state agencies should issue licenses solely to fully compliant programs. If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly, we needed to rethink our program evaluation strategies.

In the United States, state licensing and regulatory agencies establish childcare regulations, but federal agencies such as the Office of Child Care and the Administration for Children



adapted from Richard Fiene

This graph shows the quality scores (*y*-axis) associated with four categories of regulatory compliance (*x*-axis, defined by the number of rules violations, ranging from 0 [Level 1] to 10 or more [Level 4]). Note that compliance scores (*blue line*) and quality scores (*red line*) rise together, but only until substantial compliance (99-97 percent compliance with all rules (Level 2) is reached. This finding argues for the adoption of substantial compliance as a standard, and for utilizing differential/relative monitoring to better capture nuances of quality and more efficiently allocate resources. The alternative—a punitive, gatekeeping licensing approach requiring full compliance (a yes/no proposition)—has led to highly skewed data. Here, the author has split (dichotomized) these skewed data into two extremes: Programs with regulatory compliance scores in the top 5-10 percent (*upper right*, labeled KI+/RA+ to indicate positive key indicator and risk assessment findings) and the bottom 5–10 percent (*lower left*, labeled KI-/RA-). The graph shows how scores in key indicators and risk assessment effectively predict program quality.

and Families also influence rules, as does Congress through its funding purse strings. Sometimes cities and counties, too, set regulations or standards, especially concerning physical environment, health, safety, and zoning. (Here, the term "regulations" means those defined by the National Association for Regulatory Administration's Licensing Curriculum.)

For an individual program or facility to operate, a state licensing agency must judge that it follows these standards. Examples include certifications for teacher qualifications, first aid, CPR, and the facility environment, along with requirements for ongoing training and professional development. State licensing staff evaluate compliance via inspections, document reviews, audits, and interviews, usually on a yearly basis. Inspections check for health, safety, cleanliness, educational standards, and staff-to-child ratios, as well as less obvious standards such as playground and transportation safety. Noncompliant programs may face fines, mandated corrective actions, training, or technical assistance, or may undergo license suspension or even permanent closure.

Licensing requirements vary depending on the childcare offered (such as family childcare homes, center-based care, or school-based programs), with larger centers typically facing more stringent requirements. Along with compliance ratings and violations issued by licensing inspectors, these facilities voluntarily seek ratings from quality initiative offices within human services agencies.

Here, and in my research, I primarily deal with center-based care programs, but the findings apply to other service types as well, such as family childcare homes and school-age programs, as well as human services categories such as child residential, child foster care, adult residential, and adult personal care homes. My data and research concern the relationship between quality and compliance, and how to improve it. They stem from studies of hundreds of programs I conducted at the state level from the 1970s through the 2010s, when I directed various research and training institutes at Pennsylvania State University. In these controlled and replicated studies, trained observers collected both regulatory data and program quality data from eight states, three Canadian provinces, and the U.S. Head Start program. The work ran the gamut, from site selection via stratified random samples, to dispatching data collectors to specific programs, to providing individual states with an overall blueprint describing how to conduct their studies.

Initially, the ceiling effect between regulatory compliance and program quality came as a surprise; we did not predict that full compliance would fail to outperform substantial compliance. It also drew pushback from the licensing field. Thus, I replicated the study many times over to assess my assumptions. But the finding persisted: Program quality scores rise with regulatory compliance until programs reach substantial compliance, after which quality declines. Although until 1980 states required childcare programs to show full compliance and zero violations, since 2015 most states have allowed licensing for facilities that are substantially compliant.

Differential Monitoring

If substantial compliance with some rules rather than full compliance with all rules best ensures the childcare program quality, then the question naturally arises: "Which rules?" Conceivably, some rules should weigh more heavily than others—say, the ones that data show most closely relate to safety and quality. Such is precisely the idea behind *differential monitoring*.

Differential monitoring emerged in 1979 during my discussions with federal agencies such as the Administration for Children, Youth and Families and the Children's Bureau, who felt dissatisfied with the traditional uniform monitoring approach. They knew about my team's work in Pennsylvania and invited me to give a series of talks to their staff. The result was a move away from the older, one-size-fits-all approach to differential methods focused on *key indicators* and *risk assessments*.

Key indicators are statistical predictors of overall compliance—rules that, if a facility follows them, strongly suggest they will follow other rules as

well. They very efficiently determine a facility's overall regulatory compliance without requiring a comprehensive inspection. Far from negligent, this approach works because not all rules are created and monitored equally.

Risk assessment focuses on those rules and regulations which, when breached, place children at greatest risk, such as rules that deal with supervision or hazardous materials handling, among others. Generally, jurisdictions, states, and provinces engage major early care and education stakeholders (service providers, parents, advocates, and licensing staff) in weighting rules or regulations based on their risks to children's health and safety. Commonly, participants assign weights via a *Likert scale*—a common survey and questionnaire tool that lets respondents indicate the strength of their agreement or disagreement (or, in this case, their assessment of risk) with a statement about attitudes, opinions, or perceptions. The weights range from 1 to 10, where 1 indicates little risk if a program fails to follow the specific rule or regulation and 10 corresponds to high risk. Rules heavily weighted as associated with sickness, injury, or death join the risk assessment rules measured by inspectors in every differential monitoring review.

As an aside, I should point out that full compliance remains the standard for maintaining health and safety. So why incorporate risk assessments into differential monitoring and, by extension, the substantial compliance paradigm, as its own separate metric? In truth, I had no such intention when I wrote my 1985 research papers about differential monitoring and the theory of regulatory compliance. Rather, risk assessment morphed from a way to provide the needed data variance for key indicator scoring into its own submethodology. As it found its way into the implementation of national standards and guidelines, risk assessment subsequently emerged as a separate methodology.

Our findings repeatedly show that using the combined methodologies of key indicator predictor rules and risk assessment rules to identify the "right rules" and to ensure compliance with them, rather than to seek full compliance, makes the differential monitoring approach the most effective and efficient program monitoring system. Also, studies show that abbreviated, targeted, and focused reviews take approximately 50 percent less time than comprehensive reviews.

Compliance Measurement Systems

scoring level	individual rule		aggregate rules	individual rule
scale	instrument based	*scale*	differential	integrated
7	full compliance	7	full compliance	exceeds compliance
–	–	5	substantial	full compliance
–	–	3	mediocre	substantial
1	out of compliance	1	low	mediocre/low

adapted from Richard Fiene

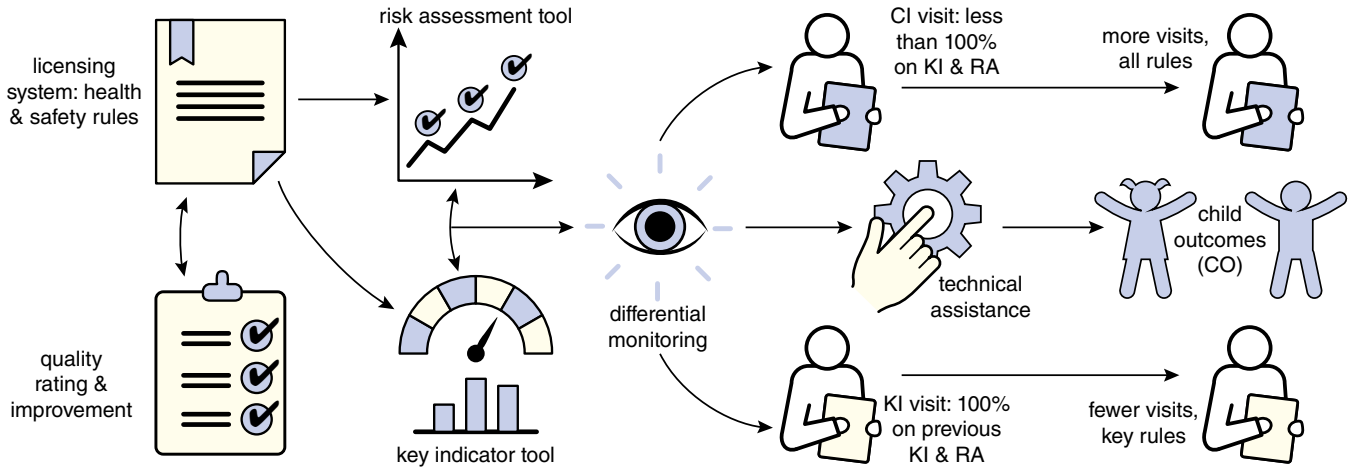
This table compares different approaches to measuring compliance: A licensing-focused approach in which programs are classified as either compliant or noncompliant based on rules violation counts, with no middle ground (*columns 1 and 2*), and a more nuanced ordinal approach using a Likert scale. This experimental metric, called the Regulatory Compliance Scale (*column 3*), is currently being tested at the aggregate rule level (*column 4*) and may be expanded to the level of individual rules (*column 5*) in the future. Note that aggregate rule scores are not equal to the sum of all individual rule scores because not all rules are created or administered equally.

Unfortunately, although many licensing bodies use risk assessment or key indicator methodologies, few use both. *Monitoring Practices Used in Child Care and Early Education Licensing*, a federal accounting of how states conduct program monitoring, reported that 10 states used key indicators, 17 states used risk assessments, and only one state used both. Hopefully, this pattern will change as the regulatory science field matures over the coming decades.

If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly we needed to rethink our program evaluation strategies.

Since I first proposed it in the mid-1980s, the theory of regulatory compliance has faced numerous critics in the human services licensing field, especially among advocates of uniform monitoring and full compliance. Only after years of licensing validation studies conducted by my team and others repeatedly demonstrated that full compliance did not produce the highest quality did states begin licensing programs in substantial rather than full regulatory compliance. Today, although various U.S. states apply the differential monitoring review approach unevenly, nearly all have adopted the policy of granting licenses for substantial rather than full compliance. The latest revision of the legislation for the Child Care and Development Block Grant (a U.S. federal funding program that helps states, territories, and tribes assist low-income families access affordable childcare) cites differential monitoring as an alternative to uniform program monitoring.

Of all the approaches and methodologies that flow from the theory of regulatory compliance, differential monitoring most significantly alters the program monitoring, inspection, review, and licensing landscape. Its reviews occur just as often as do uniform monitoring assessments but focus specifically on rule breaches shown to place children at risk. That said, differential monitoring did not replace but rather supplemented its predecessor: Comprehensive reviews must still occur every three to four years to validate the performance of key indicators and risk assessment rules. But what does that report card look like in terms of analyzable data?



Barbara Aulicino

This illustration shows the various components that contribute to a differential monitoring approach and how agencies can use them to evaluate the effectiveness and validity of different approaches. Differential monitoring allocates resources based on risk assessment (client morbidity and/or mortality) and key indicators (rules whose compliance is strongly predictive of program quality). These data, provided by mandatory licensing processes and voluntary quality rating services, reveals which programs are highly compliant with key rules (though not all rules) and therefore require fewer visits versus programs that are less compliant and require additional visits and technical assistance to achieve similar child outcomes.

Rethinking Nominal Data

Traditionally, licensing data are categorical (sorted into groups such as "approved" or "denied"), unordered (there's no built-in way for such groups to be sequenced), and mutually exclusive (state agencies cannot simultaneously deem a facility both "approved" and "denied"). In statistical terms, such data are nominal, like a table listing cars by make or model; you cannot "do math" on such a table like you can on, say, on a table listing automobile curb weights and fuel economies. It is also binary: A program either follows a rule, or it doesn't.

Presently most jurisdictions deal in these absolutes and exclude gray areas. This approach, much like uniform program monitoring and full compliance, makes intuitive sense: We create rules and regulations because we believe in the value of following them, and because licenses mean nothing if licensees are not held to a standard. But here again, we must look deeper and ask, "What consequences follow from this either/or approach to measuring compliance, and who decides whether or not a particular box gets checked?"

Let's begin with the latter question. In an ideal world, judgments made by assessors would perfectly reflect a program's actual regulatory compliance state. But research that tests reliability and replicability in the licensing field empirically shows a concerning degree of disagreement when a second observer validates the decision regarding regulatory compliance. These disagreements suggest a worrying number of false positives and false negatives.

A false positive occurs when a program follows a rule or regulation, but the assessor rules that the facility is noncompliant (which might sound backwards, but the metric is *non*compliance, not compliance, so finding a false violation means finding a false positive). But even more concerning are false negatives, in which an evaluator says a program complies with a rule that it breaches, thereby placing clients at risk. Detecting false negatives is one of the chief reasons we periodically validate the predictive value of key indicator rules through comprehensive reviews.

As for the first question, the answer is simple: Nominal, binary licensing data is severely skewed. Upon reflection, the reason becomes obvious. When a regulated industry such as childcare mandates compliance before a program can operate and excludes gray areas, most facilities will achieve full compliance or lose their licenses. Because unlicensed providers don't last long, the childcare sector produces data that skew toward licensed programs. To grasp such skewed continuous or multicategory data, we must first dichotomize it into two distinct groups.

Such sorting into piles raises statisticians' hackles; unless carefully done, it accentuates differences and forces tradeoffs between precision and sensitivity, which can mean swapping false positives for false negatives. But the nature of licensing data—a skewed collection of mostly or fully compliant programs dumped in a single bucket—makes the split both necessary and warranted. By setting a threshold of certainty or agreement among evaluators, we can more effectively reduce false negatives, that is, cases in which evaluators say a program follows a rule when it doesn't.

This need becomes even clearer when one considers the demands posed by differential monitoring and its methodologies, key indicators, and risk assessments. For a program to receive licensure, it is not enough to ask if it "complies enough overall"; we must also know if it follows the specific rules that most ensure safety. By comparing highly compliant programs only with low-compliant programs, we accentuate the differences between the two and bolster our data analyses as well as overall safety. This comports well with licensing decision-making, which can consider a program compliant or noncompliant not only in aggregate, but with respect to *individual* rules.

Infusing Quality

The all-or-nothing approach to regulatory compliance and licensing fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from a false assumption that program quality increases in step with 100 percent compliance. But I am far from the first

to notice that approach's weaknesses in evaluating how good a program or facility actually is. Indeed, its shortcomings helped drive the creation of a separate industry of voluntary accreditation programs such as the National Association for the Education of Young Children, state-run quality rating and improvement systems, and third-party tools and assessments. It's time we folded quality assessments into regulatory compliance.

I have already explained how the theory of regulatory compliance improves program quality and safety by focusing on substantial, not full, compliance and by using differential monitoring to ensure programs follow the most protective and impactful rules. But to further cast off the limitations and lopsidedness of a uniform monitoring and full compliance mindset, and to make room for data capable of tracking quality, we must also replace rigid either/or logic with a more nuanced ordinal measurement: a scaling technique.

Recall that assessors can evaluate compliance in two ways: They can consider aggregate rules—collections of rules that fall into categories such as staffing or safety practices—or individual rules. Each has its own studies and research literature. Research on aggregate rules from the 1970s, 1980s, and the 2010s established substantial compliance as a "sweet spot" of best outcomes and showed that the time had come to replace nominal metrics (such as "compliant" and "noncompliant") with ordinal ones (such as "98 percent compliant").

Inspired by this research, I have proposed replacing older nominal techniques with an ordinal scale like the Likert scale already used in quality measurements (usually but not always ranging from 1–7, with 1 being inadequate and 7 being excellent). This technique, currently under review by the National Association for Regulatory Administration, will help reviewers consider the importance of substantial compliance. Moreover, it will add the currently absent quality elements to each rule and regulation. However, this approach involves aggregate rules only; further research is needed to determine if the same shift from nominal to ordinal metrics should also occur at the individual rule level.

Should those findings bear out the value of evaluating individual rules via the 1–7 regulatory compliance scale, I propose that it should contain the following categories: exceeding full compliance, full compliance, substantial compliance, and mediocre compliance (*see figure on page 19*). These categories differ from the aggregate rule compliance scale currently under evaluation (full, substantial, mediocre, and low compliance) because aggregate compliance only considers health and safety elements, whereas an individual scale would also take quality into account.

Research supports the value of transitioning from uniform monitoring and full compliance to differential monitoring and substantial compliance. Practice has shown the value of retaining the older to help ensure the validity of the newer. Looking to the future, I believe we can further improve compliance evaluations by developing and evaluating *integrative monitoring*, which incorporates program quality into rule formulation and moves the key indicators from predicting compliance to forecasting quality.

The all-or-nothing approach fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from the false assumption that program quality increases in step with 100 percent compliance.

Looking Forward

The regulatory compliance scale is a new and evolving metric. It transforms licensing data from a mere violation tally into a more useful and intuitive scale, one more consistent with the program quality measurements supported by research. Hereafter, I hope that the approach will incorporate quality measurements and more nuanced weighting into the evaluation of individual rule compliance. But discussions are just beginning, and this shift will pose a substantial challenge for agencies, which must also cope with the aftermath of the COVID-19 pandemic and a rising tendency toward deregulation.

The theory of regulatory compliance concerns the relationship between regulatory compliance and program quality, not health and safety, where full compliance remains the goal. It is, however, the preferred methodology for eliminating false negatives and decreasing false positives. Add to that the fact that the theory of regulatory compliance predicts a nonlinear relationship between compliance and quality but a linear relationship linking regulatory compliance and safety, and regulatory scientists clearly have our work cut out for us. Untying this knot will require greater collaboration between the historically siloed public policy worlds of licensing, accreditation, quality rating and improvement systems, and professional development systems.

I hope that the regulatory science field takes these paradigm shifts into consideration as it builds licensing decision-making systems and considers how states issue licenses. And although this work deals primarily with my own experience in the early care and education field, I wonder if other human service sectors, such as the foster care or child and adult residential areas, demonstrate similar patterns. Other disciplines that deal with regulations and compliance may similarly find it fruitful to discuss the nuances of their own evaluation metrics in order to achieve the best overall outcome with the most efficient use of limited resources.

Bibliography

Fiene, R. 2023. Licensing measurement and monitoring. National Association for Regulatory Administration licensing curriculum. Published online www.naralicensing.org/key-indicators

Fiene, R. 2022. Regulatory compliance monitoring paradigms and the relationship of regulatory compliance/licensing with program quality: A policy commentary. *Journal of Regulatory Science* 10:1–7.

Fiene, R. 2019. A treatise on regulatory compliance. *Journal of Regulatory Science* 7:1–3.

Richard Fiene is an emeritus professor of psychology at the Pennsylvania State University and has had a lengthy career in regulatory science and developmental psychology, in which his research has led to the development of national early care and education standards. He directs the Research Institute for Key Indicators Data Lab. Email: RFiene@RIKInstitute.com