

A Mixed Method Program Evaluation of Annual Inspections Conducted in Childcare Programs in Washington State

Sonya Stevens^{a,*}

^a*Licensing Division, Washington Department of Children, Youth, and Families, Spokane, WA 99223*

Abstract

This mixed method study used a program evaluation to assess the reliability and social validity of the focused childcare monitoring checklist used in Washington State, as well as its social validity in maintaining quality programming in licensed childcare centers. The focused monitoring checklist and interview responses were used to answer two specific research questions: (1) How do stakeholders describe the value, usefulness, and effects of state administrated focused monitoring?; and (2) What is the inter-rater reliability of the focused monitoring tool used to assess the foundational health and safety issues that must be met by state licensed early childhood programs? The study found that licensors and providers found the focused monitoring tool as more efficient and informative than the current differential monitoring system. The use of a checklist focusing on real time compliance increased the value placed on the relevance of the inspection with respect to meeting licensor and provider needs. The results also showed that even with a controlled tool, performance of onsite inspections can vary greatly along a continuum of reliability and objectivity due to licensor rater drift and individual perceptions of licensing procedures. Licensing agencies should consider further evaluation of the monitoring process and the reliability of the checklist tool as the process is implemented statewide, concentrating on the training content and training methods provided to licensors.

Keywords: licensing, regulatory oversight, childcare, inter-rater reliability, differential monitoring

1. Introduction

According to the U.S. Bureau of Labor Statistics, 72 percent of children nationwide were cared for in non-parental settings beginning in the first year of life in 2013 [13]. Although the direct responsibility for ensuring the health and safety of children in licensed childcare lies with families and local childcare providers, most states have recognized the importance of providing oversight and support to childcare providers through formal inspections [16, 19]. Early childhood centers are required by state licensing agencies to ensure that young children are safe, healthy and ready to learn [6]. Ensuring the use of health and safety practices in early childhood centers requires formal oversight via a monitoring system.

Agencies across the United States are working to ensure providers meet overall health and safety standards, as well as improved quality, through effective and reliable monitoring to ensure compliance with regulations [1, 28], but the reliability and social validity of the monitoring tool to assess the quality of care for children has yet to be considered. Wolf [29] described social validity as the need to allow stakeholders to val-

idate critical work to ensure it meets their needs and the results of the work justify the means to get there. Thus, the social validity of a state-administrated monitoring system can be measured by inviting stakeholders to participate in the evaluation of the value, usefulness, and effects of a state-administrated monitoring system on maintaining quality programming and increasing program outcomes for children [26, 20].

One example of a state-administered monitoring system is differential monitoring. A differential monitoring system can be used to recognize a provider's strong record of compliance by using abbreviated, or less frequent, inspections [7]. Differential monitoring was employed in Washington State in 2012 with the intent to save licensors time, thereby saving the licensing agency's money by concentrating on sites with significant compliance issues, rather than the entire set of regulations. This methodology identifies those programs with a history of high compliance allowing the use of an abbreviated checklist [8]. Differential monitoring was believed to allow licensors to concentrate more heavily on current areas of non-compliance, while limiting time spent on areas where providers need minimal or no licensor support during any given inspection.

The research study outlined in this article was used to evaluate a differential monitoring methodology when combining the

*Corresponding author: Sonya Stevens, Email: mrssonya.stevens@gmail.com

abbreviated checklist and a comprehensive checklist into one checklist system. Differentiation of the checklist within this study is defined as an abbreviated inspection checklist for all providers, which expands in real time based on findings during the monitoring visit, rather than based on a historical level of compliance.

An annual monitoring tool that is reliable and socially valid could potentially provide the support needed to increase compliance with all licensing rules, ensuring foundational levels of quality, safety, and health, as well as inform the agency on how to provide licensing consistency that might lead to increases in liability protection [16]. Two research questions were addressed: (1) How do stakeholders describe the value, usefulness, and effects of state-administrated focused monitoring?; and (2) What is the inter-rater reliability of the focused monitoring observation tool used to assess the foundational health and safety concerns that must be met by state-licensed early childhood programs?

The purpose of the mixed-method program evaluation was to determine the reliability of the focused monitoring tool and the social validity of the focused monitoring processes used to monitor the foundational health and safety of childcare programs in Washington State. The problem addressed in the present study is that the reliability of the monitoring tools and the social validity of the monitoring process used to assess annual compliance of licensed childcare centers has not been determined, which calls into question a state agency's ability to ensure a foundational level of health and safety for children in licensed childcare.

2. Background

When children are cared for in environments that are safe, healthy, and developmentally appropriate, they experience fewer illnesses and injuries, are more likely to have better health care, receive accurate referrals to developmental services when needed, and are simply more prepared and ready to learn. The process of evaluating licensed childcare remains critical because children are a vulnerable population and often do not have the communication skills or knowledge to sufficiently advocate for their right to a safe, healthy, and developmentally appropriate care environment [19, 16, 23]. It is for these reasons that there have been multiple state and federal initiatives to improve health and safety for children in childcare settings [6, 21].

Perhaps the largest initiative to increase health and safety in childcare comes from standards outlined in the publication *Caring for Our Children* (CFOC) [6, 14]. Crowley et al. [6] described CFOC as an inclusive set of standards developed in partnership by the U.S. Department of Health and Human Services and the American Academy of Pediatrics that is aligned with accreditation standards of the National Association for the Education of Young Children (NAEYC). CFOC presents common health and safety standards that can reduce conflicting standards sometimes found when considering multiple funding [6, 14].

Agencies should be careful to avoid making the regulatory design overly complicated and strict. Moloney [23] found that teachers in England reported some increase in concern around inspector inconsistencies and their lack of understanding of the needs of early learners. Because of this lack of confidence toward the inspectors, the importance of the inspection and its findings was not considered relevant by those being inspected [23]. In order to increase confidence toward the inspector and the inspection, state agencies must continuously assess monitoring systems to include holistic programming support and work to develop systems that not only make sense to a regulatory agency, but can also be a useful tool for regulatory improvement and consistency.

All states require routine inspections to monitor compliance of state-mandated childcare regulations, but no two systems are alike [1, 6]. Much of the research completed over the past three decades defines three key theoretical approaches to monitoring licensed childcare for improved health and safety in group care, as well as limited requirements addressing social and emotional development [8, 10, 11]. The three theoretical approaches include routine unannounced inspections; a combination of standard weighting (the classification of regulation importance) and/or a key indicator system (standards with statistical evidence of reliability or probability of full compliance or non-compliance); and differential monitoring (the frequency and depth of inspections based on a provider's history of compliance with regulations) [8, 6]. Most states use either a weighted system of monitoring or an indicator system, but few employ both [8].

Licensing key indicators are defined as specific rules statistically predicting overall compliance with all the licensing rules [10]. It is important to note that key indicators are not necessarily those rules most frequently found out of compliance or that place clients at the greatest risk. Abbreviated inspection includes key indicators, and the methodology outlines that if one of the indicators within the key categories is found to be non-compliant, then it is probable that other areas within the program will also be out of compliance, which would require a holistic look at full compliance [10]. In other words, programs historically in high compliance would be able to have an abbreviated inspection, and if a key indicator was out of compliance, the inspection would have to be switched to a comprehensive rule inspection.

Washington State used a combination of both the key indicator system and the differential monitoring system [24, 15] beginning in 2012. Fiene's [10] 13 key indicators are used as the core of an abbreviated checklist used for monitoring visits by providers with high levels of compliance [24]. Being found in non-compliance on one of the key indicators triggers the licensors to move to a comprehensive checklist, making the system both a differential and key indicator system [28, 24, 15]. A 2014 assessment of the key indicators and differential system found that from a sample of 1,401 monitoring visits conducted in 2014, approximately 11 percent of the total licensing visits did not switch from an abbreviated checklist to a comprehensive one when required. The fact that a significant amount of monitoring visits did not follow outlined methods of the moni-

toring model, and the fact that research has not demonstrated a level of social acceptance of the monitoring system, made the timing of this research relevant, as many states are assessing their own monitoring processes in child care licensing.

3. Population and Sample

The Washington State Child Care licensing division is separated into four different regions. Region 1 (Eastern Region) is the largest geographically, covering 20 counties east of the Cascade Mountains. This region provides oversight from four licensing offices located in the cities of Yakima, Wenatchee, Tri-Cities, and Spokane, and is responsible for approximately 1,548 childcare providers. Region 2 (Northwest Region) contains five northwestern counties in the state and provides oversight from four offices located in Bellingham, Mount Vernon, Everett, and Bellevue. Region 2 is responsible for approximately 1,412 licensed providers. Region 3 (Southwest Region) includes 13 counties northwest of the San Juan Islands down to the Oregon border and has about 1,431 licensed providers. Region 3 has three main licensing offices in Olympia, Tacoma, and Vancouver, and three satellite offices in Aberdeen, Kelso and Port Angeles. Region 4 (South King County) provides oversight from three licensing offices, all located in Renton, with approximately 1,366 licensees (DEL.wa.gov, 2016).

The Department of Early Learning (DEL), now the Department of Children, Youth, and Families (DCYF), employs approximately 85 licensors to conduct annual site visits across the state. To be a child care licensor in the state of Washington, applicants must have, at minimum, one-year experience as a social worker or a master's degree in social services, human services, behavioral sciences, or a related field, as well as related work experience. A licensor may be hired if he/she has a bachelor's degree in those same fields of study and three years of equivalent experience. Licensors are expected to have knowledge of child development, risk assessment, interview techniques, service delivery, and community resources, among other subjects. Childcare licensors in Washington State vary in age, gender, and ethnicity, and all are fluent in English.

An invitation to participate in the present study was sent through the agency communications office to all childcare licensors in Washington State. Any licensor who wanted to, and had the approval of their licensing supervisor, was allowed to participate. Interested participants were invited to contact the researcher directly, to ask questions before agreeing to participate. Licensors had up to one week to consider their decision. In the end, seven ($n=7$) licensors were recruited and selected, representing each of the four regions of Washington State. Two licensors were stationed in the Eastern Region, one in the Southwest Region, one in South King County, and three in the Northwest Region.

Provider sites were recruited by searching through the licensor Toolkit (the agency provider database), to identify those sites with annual monitoring visits due during the study timeline, to ensure that all the monitoring visits would fall within the legal timeline outlined by Washington State law. From

this population, a random sample of 245 providers were contacted through a general email delivered by the DCYF communications office, requesting volunteer participants. Participants were not offered any compensation or assurances that participation would not change the results of the monitoring visit. Once site directors or program owners indicated a willingness to participate, the researcher contacted the site administrator, provided details as to the purpose of the research, explained what they would be asked to do if they decided to participate in the project, and answered any questions. As a result, five ($n=5$) center providers agreed to participate in the study. One provider was located in the Eastern Region of Washington State, three providers were located throughout the Southwest Region, and one provider was in the Northwest Region.

Center owners or directors consenting to participate were asked to be both the primary contact for the licensing monitoring visit and the individual working with the licensor during the actual monitoring visit. At one of the centers, there was an additional staff person on site who was responsible for working with the licensor during the site visit. That individual was also given the same consent information provided to the center director or owner, which resulted in six ($n=6$) provider participants at the five participating centers. Potential provider participants were also given two weeks to consider their decision and return the informed consent form. Consent specifically outlined the ability to schedule interviews at the convenience of the participant, but did not specify the date or timeline of their annual monitoring visit, as those remained unannounced. Provider interviews were conducted in the provider participant's place of business, so as to not require participants to travel or unduly interrupt daily operations.

4. Methods

A program evaluation research design is used to guide the decision-making process concerning program policy and/or management when the purpose of the research is to make judgments about the merits or value of an educational program [12]. Specifically, the Context, Input, Process, and Product Evaluation (CIPP) [27] was used in the present study as the framework for integrating program evaluation with program management and policy development. CIPP involves the performance of three broad tasks for each type of evaluation (i.e., context, input, process, product) in the following order: (1) delineate the information needed for decision making; (2) obtain the information; and (3) synthesize the information to make programmatic decisions.

4.1. Context

Context evaluation involved the identification of specific problems regarding the inspection checklist's social validity over the last few years, and provided the basis for developing objectives that will lead to program improvement. To delineate the information needed for decision making, a committee of in-

ternal Department of Early Learning¹ (DEL) licensing staff met during three one-hour webinar meetings. The researcher shared thematic information from the literature review and other internal documentation with the committee, so that the committee could develop objectives they felt would lead to program improvements.

4.2. Input

Input evaluation involved collecting information about the resources and outlining strategies needed to accomplish program goals, including the determination of constraints. The aforementioned ad-hoc committee collaborated on three additional webinar meetings to identify contents and structure of the focused monitoring tool. Washington Administrative Codes (WACs) and the Revised Codes of Washington (RCW) were reviewed, for they are the core content within the focused monitoring checklist. Guidance for the licensors use of the focused monitoring tool were also discussed. Overall, the committee worked to develop the focused monitoring instrument, monitoring system, and training outline.

4.3. Process

The process evaluation collected data to monitor the program as it operated day to day, which allowed the identification of problems in their early stages. This included training to prepare participating licensors to implement the focused monitoring process during site visits. In order to delineate process data, licensors were paired together to conduct monitoring visits simultaneously, using identical focused monitoring checklists. These data determined if an adequate level of inter-rater reliability had been met. Data were obtained through a three-week long pilot study using the redesigned checklist, followed by in-person interviews to gather data for the formative and summative assessments.

4.4. Product

The product evaluation utilized data to determine the extent to which the goals of the program had been achieved, and to make decisions about continuing or modifying the project. The results were delineated to the original committee in a final one-hour webinar meeting. The committee made recommendations for additional programmatic modifications to the focused monitoring system and the tools used to conduct annual inspections.

5. Data Collection, Processing and Analysis

5.1. Context and Input

All field notes, meeting transcriptions, and interviews were sorted into separate databases by phase (context, input, and product) and entered into a coding analysis software program (NVivoTM) file for coding and analysis. Data were coded and re-coded into common themes and used to help determine the focused checklist developmental needs, as well as the training protocols for the volunteer licensor participants.

¹In 2018, DEL became the Department of Children, Youth, and Families (DCYF).

5.2. Process

Licensors used the focused monitoring checklist developed during the input and context phases to collect data during the process evaluation, using a pencil and paper version of the checklist during the unannounced inspection of each participating childcare center. Two licensors were assigned to each site visit. One licensor served as the primary inspector and was the lead on communicating the results to the provider, writing licensor notes, and completing the paperwork. The second licensor completed an identical checklist simultaneously, only for the purpose of evaluating inter-rater reliability. The identification of which licensor was the primary and which licensor was the secondary was kept from the licensors until after the checklists were complete. The primary licensor was then responsible for communication and all follow up with the provider. Finally, the licensors were asked to not communicate with one another regarding the visit details during the licensing visit to ensure independent assessments of the provider.

Data collected from the paper copies of each of the focused monitoring checklist were compiled and coded into an Excel database. Standards marked compliant were coded as 2, non-complaint coded as 1, and discussed was coded as 0. Inter-rater reliability was determined using a simple computed formula, dividing the number of times two licensors agreed (indicated by a +) on the checklist during the same monitoring visit, by the sum total items on the checklist. An acceptable level of inter-rater reliability was no less than 85 percent degree of inter-rater reliability [1, 5].

Data were gathered about the social validity of the newly developed process by conducting post-inspection interviews with the seven participating licensors and six providers. All interviews except one (provider did not consent to audiotaping) were taped and transcribed verbatim, de-identified, and entered into an NVivoTM file for analysis. NVivoTM organized the interview data and analyzed them to identify emergent themes based on questions designed to address specific areas of social validity by type of interview.

5.3. Product

The results of the context, input, and process evaluations were summarized and reported to the committee. The researcher and the committee discussed necessary revisions to the training materials and the focused monitoring checklist in preparation for a final report and recommendations to be submitted to the agency leadership. Additionally, committee participants were asked similar questions to those asked during the licensor and provider interviews to further evaluate the topic of social validity. The final meeting was recorded, and, once again, transcribed verbatim and put into NVivoTM coding software for analysis.

6. Results

The purpose of this study was to evaluate how stakeholders describe the value, usefulness, and effects of state-administered focused monitoring, and to assess the reliability of the focused

Theme	Total Count Field Notes References	Total Count Historical Data References	Total Count Literature References	Total
Compliance	2	3	15	20
Differential Monitoring	1	3	13	17
Current Checklist Challenges	15	0	0	15
Focused Monitoring	1	6	5	12
Resources Needed	8	3	1	12
Checklist Use	3	5	0	8
Quality Improvement	0	0	7	7
Safety and Health	1	2	3	6
Non Compliance	2	2	2	6
Scoring System	3	0	2	5
Weighted Regulation	0	0	2	2

Table 1: Context Data Collection Themes and Frequency Count by Ad-Hoc Committee.

monitoring tool licensors use to conduct annual inspections in childcare facilities in Washington State. The CIPP process was used to guide the evaluation, whereby three broad tasks were conducted for each type of evaluation in the following order: (1) delineate the information needed for decision making; (2) obtain the information; and (3) synthesize the information to make programmatic decisions.

6.1. Context Evaluation

The context evaluation was used to evaluate and redefine the current checklist system in order to guide the development of a checklist methodology for this study. Through the context evaluation process, the committee was able to understand how the current checklist system was being utilized by licensing staff and where programmatic improvements may be needed. The information was obtained by collecting and coding all literature data, including previous focus groups notes, legislative reports and leadership meeting notes. Codes were used to identify themes within the narrative context. Table 1 shows the themes found in the context evaluation.

The themes with the highest frequency included compliance, differential monitoring, current checklist challenges, focused monitoring, and resources needed. From this information, the committee agreed the program evaluation goals would: 1) create a differential monitoring system that, through improved compliance by addressing the current checklist challenges, would improve the health, safety, and development of children in licensed childcare; and 2) create a focused checklist that provides consistency and is purposeful and meaningful to licensors and providers in relation to all regulations.

6.2. Input Evaluation

The information for the input evaluation was designed to understand specific problematic areas that needed to be corrected in the current checklist before moving forward with the focused checklist design. The information was obtained in an ad-hoc committee meeting by reviewing context data and identifying areas needing specific changes. Collected narrative data were synthesized into themes. Table 2 shows the themes found in the context evaluation.

The ad-hoc committee identified seven broad issues needing attention within the current checklist challenges (CCC) theme: (1) redundancy when moving from the abbreviated checklist to the comprehensive checklist; (2) not enough information on the abbreviated checklist; (3) abbreviated checklist is always the same regardless of the provider type or need; (4) not enough clarity regarding the rule; (5) inconsistency by the licensors, including inconsistent use of the checklist; (6) items on the checklist that do not apply to the specific provider; and (7) excessive use of adding non-compliances that were not already on the checklist. The ad-hoc committee also made five general suggestions for improvements: building a checklist that is meaningful to each provider; ensuring a comprehensive baseline by rotating random items over a three-year time frame; reducing redundant items and eliminating items that do not apply; providing direct resources specific to each provider; and including the weights to ensure more consistent citations.

6.3. Process Evaluation

Data were collected and analyzed from seven licensor interviews and six provider participant interviews to conduct the

Theme	Total Count Checklist Development	Total Count Training Development	Total
Current Checklist Development	9	9	18
Checklist Development Challenges	5	3	8
Pilot Training Needs	2	5	7
Checklist as a Resource	6	1	4
Monitoring Consequences	4	0	4
Future Training Needs	0	2	2

Table 2: Input Data Collection - Theme and Frequency Mentions by Ad-Hoc Committee Members.

Theme	Total # of Licensor Mentions	Total # of Provider Mentions	Total
Informs Program Needs	25	17	42
Effect Quality	19	13	32
Challenges With the Current Checklist	14	11	25
Other Resources	14	10	24
Checklist Opportunities	10	9	19
Process Easier	13	3	16
Process More Difficult	11	5	16
Monitoring Rewards	10	6	16
Checklist as a Resource	8	5	13
Process Not Different	6	6	12
Reduce Workload	1	9	10
Training Needs	10	N/A ^a	10
Not Effect Quality	6	3	9
Training Useful	8	N/A	8
Monitoring Consequences	4	1	5
Not Informed Program Needs	3	1	4
Training Not Useful	2	N/A	2
No Workload Change	1	1	2

Table 3: Process Evaluation Themes Identified by Licensor and Provider Participants.

^aDenotes questions asked of only one set of participants. For example, TU (training useful) was only asked of licensor participants.

process evaluation. The data were used to inform a final programmatic decision within the product evaluation. Table 3 shows the themes that emerged as a result of the process evaluation.

Licensors referenced the challenges with the current checklist system 25 times, while providers referenced it 17 times, for a total of 42 references. The concern most frequently mentioned by both licensors and providers was that items being inspected on the current checklist often did not apply to the provider being inspected. For example, one of the participating licensors observed that, “if you follow the short checklist that we currently use, you are going to miss some very important critical health and safety issues . . . With the long checklist then you get into so much stuff that you don’t need . . . that it becomes very punitive rather than helpful”. The licensor and provider participants also found that items on the checklist are lacking detail, leaving room for interpretation and unnecessary variation in the checklist use. For example, one licensor stated, “With the short checklist, it’s so short, these are the things that are supposed to be important but there are other things we should be focusing on”. Providers also noted that licensor judgement played a role in the items observed and the results of the evaluation. According to one provider, “And sometimes whenever you get licensors, whatever their background is affects how they see your daily routine”.

Critical to answering the first research question: ‘How do stakeholders describe the value, usefulness, and effects of state administrated focused monitoring?’, was the understanding of how licensors and providers described the value, use and usefulness of the tool. While looking at the themes of how the focused monitoring process informed program needs and affected quality, there were 74 combined responses. Juxtaposed to this, there were far fewer responses regarding not informing program needs and not affecting quality, with only 13 total references. These responses mentioned how the design of the focused monitoring checklist can affect program needs and quality improvement, and where it still needs improvement. For example, a provider stated, “I do believe in general that it’s a good idea to focus on if we struggle with something then help give us the support around that”.

All seven licensors reported a potential to increase the time for in-depth conversation with the provider in those areas of needed improvement by eliminating the need to review irrelevant and redundant information with the provider and focusing only on expanded sections where indicators were found non-compliant. One licensor stated, “If I go in and they are over capacity then yeah, I should focus on that, on staff, on environment, on capacity issues . . . pets in their handbook, or their pools; I shouldn’t have to worry about that”. Using the random regulations rather than all the regulations would allow licensors to focus on just those areas needing support.

The licensors explained that by focusing on the provider’s needs, the checklist would not only prove to be a useful resource as they prepare for licensing visits, but would help reduce the workload (or not hinder workload). Two licensors, however, stated that it would not change workload. These two licensors also reported that they chose not to utilize the differ-

ential monitoring checklist and directions, stating that they preferred to inspect all regulations during each monitoring visit. One of these licensors stated, “I am still going to look at all of the areas regardless and cite them if there are other violations regardless”. The statement indicated that even though instructions were provided that only items showing up on the checklist could be marked non-compliant, they were willing to diverge from the outlined practice to monitor in the way they wanted.

6.4. Product Evaluation

The product evaluation was conducted during the final meeting of the ad-hoc committee. The codes and frequency from the previous evaluations were used to identify the level to which the study met its goals, and to make further adjustments and recommendations for focused monitoring in Washington State. In conjunction with the input from licensors and providers, the committee was able to explore and discuss the social validity of the focused monitoring checklist. Table 3 identifies the themes that emerged as a result of that evaluation.

Current opportunities were referenced 10 times. It was determined that because licensors did not use their usual inspection routines, they were able to pay closer attention to the items on the checklist [18]. There was also consistent conversation regarding the tool’s ability to inform program needs, provide monitoring rewards, and affect quality. Recommendations included making recommendations for changes to the regulations for clarity and common understanding, expanding on the policies and procedures during training, and further exploration on how an automated checklist system would eliminate possible user errors and increase the tool’s validity.

6.5. Inter-Rater Reliability

The results of the inter-rater reliability analysis are presented in Table 4. Industry standards of acceptable levels of inter-rater reliability vary greatly but are often considered to be between 80-85 percent and above [2, 4]. The results of the focused monitoring site visit were between 94 and 67 percent, with two site visits falling within the acceptable range. Variations in the score were evaluated by comparing the various coded data as well as narrative data already discussed.

In the high range of inter-reliability, Site A, located in the Eastern Region, scored 94 percent inter-rater reliability and was completed by two licensors from the Eastern Region, with only two issues found non-compliant. During their interviews, both licensors expressed having followed the protocols for the study. Also at the high end of inter-rater reliability, Site E, located in the Southwest Region, was completed by one licensor from the Southwest Region and one from Eastern Region, with a score of 84 percent.

On the lower range of inter-rater reliability, Site C, located in the Southwest Region, was completed by licensors from the Eastern and Southwest Regions. The inter-rater reliability score was 70 percent, with 14 findings of non-compliance. Site D, in the Northwest Region, was completed by two licensors from the Northwest Region, with the lowest score, 67 percent, and the highest number of non-compliance findings: 41. Both licensors

Site #	Licensors ID #	Inter-rater Reliability
Site A	L1006/L1007	94%
Site B	L1009/L1012	79%
Site C	L1006/L1008	70%
Site D	L1010/L1011	67%
Site E	L1006/L1008	84%

Table 4: Inter-Rater Reliability Results by Site and Rater Participation Identification.

Provider Site	Licensors Participation #	Provider Participation #	# of Historical Non-Compliant Items	# of Study Non-Compliant Items
A ^a	L1006/L1007 ^b	P1001	1	2
B	L1009/L1012	P1002/P1022	5	22
C	L1006/L1008	P1003	0	14
D	L1010/L1011	P1004	13	41
E	L1006/L1008	P1005	12	7

Table 5: Historical Number of Non-Compliance Items Inclusive of a Three-Year History vs. Number of Non-Compliance Items Found During the Study.

^aProvider sites are represented by letters A-E.^bLicensors and providers that completed the inspection are represented by their assigned participation number.

expressed a personal inspection strategy of using all standards and disregarding the differential monitoring strategy during this visit, stating that they always inspect using all regulations.

Finally, as a measure to consider if the number of items found non-compliant could be related to the inter-rater reliability, the quantity of items not in compliance on each provider's previous three years were compiled and cross referenced with the quantity of items found non-compliant from each focused monitoring visit. These data compared the combined findings from the previous three years of monitoring visits to the number of findings from the focused monitoring visit. The results are presented in Table 5. These findings show that two sites had significantly more standards found out of compliance than in the previous three years. These data were compared to the narrative data to identify if the perception of the site visits' success can be compared to the amount of issues out of compliance. It was found that within the monitoring visits where licensors expressed an unwillingness to follow instructions, the amount of non-compliances were higher.

Table 5 indicates that the inspections using the focused monitoring checklist resulted in a significant increase in findings for three of the provider sites. Notably, the provider sites with higher number of findings were also those with lower inter-rater reliability scores and lower satisfaction with the process. Site D received 41 findings of non-compliance, and, when

asked if program needs were met, provider P1004 stated "No, not in our situation". Likewise, licensor L1010 inspecting that site stated, "I feel like there are more that should be looked at on a regular basis". Site B received 22 findings of non-compliances, and when asked about affecting program quality, provider P1022 stated, "They let us know where we were deficient. Other than that, I don't see that much of a difference".

Estimating the overall inter-rater reliability alone is not an adequate assessment of inter-rater reliability. Within this study, the varying degrees of inter-rater reliability demonstrated several points of licensor accountability needing to be addressed. For example, the results from Site D demonstrated a score below 70 percent inter-rater reliability, while results from Site A were high at 94 percent. As mentioned earlier, licensors inspecting Site D made comments indicating a mistrust of the indicator methodology such as, "because when I do a monitoring visit I look at everything no matter what checklist I am using". Conversely, licensor L1006, explained to the site director, "I told her that it wasn't on my list so I am not going to look at your drills". This statement indicated the licensor understood the indicator methodology. Overall, the data demonstrated that those who had low reliability scores were those who found the most incidents of noncompliance, and were also those who reported they did not use the focused monitoring checklist. Additionally, satisfaction with the focused monitoring tool was

higher with those inspections that yielded fewer findings, while indifference to the tool itself was most prevalent from those licensors and providers with more non-compliance citations.

7. Discussion

Overall, licensors and providers described the focused monitoring checklist as an improvement to the current differential monitoring system. In general, licensors and providers believed focused monitoring would provide the information and potential resources needed for program improvement with improved usability of the checklist system itself. On the other hand, the way in which the focused monitoring tool was used in relation to fairness, objectivity, and consistency needs further exploration.

Perhaps the biggest concern that surfaced was the fact that the use of the differential checklist system varied greatly between regional users. Some licensors indicated that they are not doing a good job of regulating if they do not look at everything. However, it is impossible to have a differential system if licensors continue to look at all regulations during every monitoring visit, as supported by Fiene [8, 10, 11]. This practice is contrary to the methodology and the development of a reliable checklist if not used consistently across regions. The results of the study indicated that the accurate use of the checklist results in higher levels of confidence of the focused monitoring tool to successfully inform program needs and quality improvement. These findings suggest that there is a profound need for expanded training on specific agency goals, ideals, and culture, as well as continued assessment of inter-rater reliability. As a general limitation of this study, and studies like it, the sampling of the participants was limited in size and representative of only Washington State. Providers and licensor's in other states are monitored differently and may have different outcomes when evaluating oversight systems [22, 25].

The larger struggle of this evaluation process was the identification of large variations in inter-rater reliability. If a differential monitoring model is used by the agency for annual monitoring inspections, then appropriate use of the tool should be a clear expectation of the agency, in order to ensure the reliability of the checklist. Three of the seven licensor participants indicated a mistrust of the key indicator system, and a disregard for monitoring policy. Because child care licensing had been using differential monitoring and a key indicator system since 2012, this finding was unexpected and highlighted the fact that without a clear understanding of the key indicator system, licensors will continue to employ any checklist tool differently, resulting in the same problems being transferred to the focused monitoring system.

Estimating the overall inter-rater reliability alone is not an adequate assessment of inter-rater reliability. Within this study, the varying degrees of inter-rater reliability demonstrated several points of licensor accountability needing to be addressed. The wide variation between the inter-rater reliability findings demonstrated a need to look closer at training needs [18]. While the current study concentrated on the use of the tool, it also

demonstrated that additional consistent training is needed regarding the overall process used across licensing regions. Particularly, an emphasis on the use of differential monitoring and an understanding of the key indicator methodology is needed to reduce the variation in the use of any checklist system. To ensure a reliable focused monitoring checklist system, licensors must learn to use the tool to inform programs of items needed to maintain children's safety while in licensed child-care. Licensors must be taught to use the tool to only inspect those items identified by the differential system, introducing the focused system over time using clearly defined regulations, as supported by Kelley [17].

The present study was conducted to investigate the use and usability of a focused monitoring checklist to conduct program inspections in five childcare facilities in Washington State. Thus, the results cannot be generalized to any other state childcare facilities, family home childcare programs, or any other licensors or providers who did not participate in the study. Providers and licensors in other states are monitored differently and may have different outcomes when evaluating oversight systems [25, 22].

The present study extends the literature of methods used to measure childcare compliance by including how those effected most by the monitoring system (licensors and providers) describe its use and usability to affect positive outcomes for childcare programs and the children they serve. The foundational work presented here could provide a framework to guide other licensed oversight systems, including, but not limited to, school age, adult care programs and child welfare in Washington State and throughout North America. Additional research should be conducted in those environments to increase the generalization of the findings to all provider types.

Inter-rater reliability is an integral part of an observational tool's usefulness in the assessment of program compliance. The present study demonstrated that inspectors who do not understand or trust the tool may alter its use to meet their needs. In order to increase a monitoring checklists social validity, licensing agencies should consider soliciting further evaluation research concentrating on the training content and training methods provided to licensors. Furthermore, to ensure inspection methods are correctly and reliably implemented, special consideration should be taken to employ appropriate and consistent supervisory methods and intermittent reliability checks to ensure checklists are used correctly.

8. Conclusion

If providers do not trust the outcomes of monitoring inspections as a valuable and useful tool for improvement, regulatory outcomes are less likely to be important to the provider. Additionally, if providers have limited investment and communication with the licensing agency, the latter may hold negative views about the provider community. There is limited empirical research assessing the monitoring systems used to provide licensed childcare oversight. The little evidence that does exist suggests the content of the checklist tools must not only be

reliable but also have sufficient content validity to be successful. This study demonstrated how those most affected by licensing inspections must be encouraged and empowered to use the checklist and its procedures as a resource to maximize program improvement, as well as strategic planning for maximization of health and safety within licensed childcare programs.

9. Declaration of Conflicting Interest

The author declares no conflicts of interest.

10. Disclaimer

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

11. Acknowledgement

I would like to express my sincere gratitude to my dissertation chair Dr. Elizabeth Hardman and to the leadership of the Department of Children, Youth, and Families including but not limited to Director Dr. Betty Hyde and Director Ross Hunter, Assistant Director Luba Bezborodnikova, and Ann Radcliffe.

12. Article Information

This article was received January 2, 2020, in revised form March 25, 2020, and made available online July 19, 2020.

13. References

- Alkon, A., Rose, R., Wolff, M., Kotch, J. B., & Aronson, S. S. (2015). Health and Safety Checklist for Early Care and Education Programs to Assess Key National Health and Safety Standards. *Maternal and Child Health Journal*, 20, 114-127. doi: 10.1007/s10995-015-1809-8
- Bradshaw, C. P., Milam, A. J., Furr-Holden, C. D. M., & Lindstrom-Johnson, S. (2015). The School Assessment for Environmental Typology (SAfETy): An Observational Measure of the School Environment. *American Journal of Community Psychology*, 56(3-4), 280-292. doi: 10.1007/s10464-015-9743-x
- Bromer, J., McCabe, L. A., & Porter, T. (2013). Special section on understanding and improving quality in family childcare: Introduction and commentary. *Early Childhood Research Quarterly*, 28(4), 875-878. doi: 10.1016/j.ecresq.2013.08.003
- Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly*, 20(3), 345-360. doi: 10.1016/j.ecresq.2005.07.005
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart, and Winston.
- Crowley, A. A., Sangchoon, J., & Rosenthal, M. S. (2013). Health and Safety of Child Care Centers: An Analysis of Licensing Specialists' Reports of Routine, Unannounced Inspections. *American Journal of Public Health*, 103(10), e52-e58. doi: 10.2105/AJPH.2013.301298
- Fiene, R. (2013). *Differential monitoring logic model and algorithm (DMLMA)®: A new early childhood program quality indicator model (ECPQIM4)® for early care and education regulatory agencies*. Research Institute for Key Indicators. Retrieved from <http://www.researchconnections.org/childcare/resources/24950/pdf>
- Fiene, R. (2014). *Georgia childcare licensing study: Validating the core rule differential monitoring system*. Bright From the Start. Retrieved from <http://www.researchconnections.org/childcare/resources/30431>
- Fiene, R. (n.d.). *The relationship of accreditation and licensing standards, training, and child development outcomes*. Pennsylvania State University Capital Area Early Childhood Training Institute. Retrieved from <https://www.researchconnections.org/childcare/resources/9067>
- Fiene, R., & Carl, B. (2010). 13 Key Indicators of Child Care quality: Child Care Quality Indicators (CCQ1 – CDPES2 PC1) Scale. Retrieved from <https://harrisburg.psu.edu/files/pdf/371/2017/02/20/13keyindicatorsofchildcarequalitychildcarequalityindicatorsccqi-cdpes2pc1scale.pdf>
- Fiene, R., & Nixon, M. (1985). The Instrument Based Program Monitoring Information System and the Indicator Checklist for child care. *Child-care Quarterly*, 14(3), 198-214. doi: 10.1007/BF01113748
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction, 8th Edition*. Pearson.
- Huston, A. C., Bobbitt, K. C., & Bentley, A. (2015). Time spent in childcare: How and why does it affect social development? *Developmental Psychology*, 51(5), 621-634. doi: 10.1037/a0038951
- Hudson, G., Miller, G. G., & Seikel, K. (2014). Regulations, Policies, and Guidelines Addressing Environmental Exposures in Early Learning Environments: A Review. *Journal of Environmental Health*, 76(7), 24-34.
- Hyde, B. (2011). Childcare licensing in Washington: A report to the Washington State Legislature. Washington State Department of Early Learning. Retrieved from https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=LicensingProvisoReport11511_6bd0983a-3c05-4657-a8ba-5c6476f0d19f.pdf
- Kayira, T. (2016). Fundamentals for effective monitoring of child-care institutions. *Commonwealth Law Bulletin*, 42(1), 105-124. doi: 10.1080/03050718.2016.1143380
- Kelley, K. S., & Littenberg, B. (2015). Inter-rater reliability for scoring children's dichotic words test responses. *International Journal of Audiology*, 54(10), 751-757. doi: 10.3109/14992027.2015.1052103
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and Reliability of Measurement Instruments Used in Research. *American Journal of Health-System Pharmacy*, 65(23), 2276-2284. doi: 10.2146/ajhp070364
- Lapp-Payne, A., & Testa, M. S. (2014). *Protecting Children's Health and Safety in Every Setting*. Exchange Press.
- Ledford, J. R., Hall, E., Conder, E., & Lane, J. D. (2016). Research for young children with autism spectrum disorders: Evidence of social and ecological validity. *Topics in Early Childhood Special Education*, 35(4), 223-233. doi: 10.1177/0271121415585956
- Meloy, M. E., Lipscomb, S. T., & Baron, M. J. (2015). Linking state child care and child welfare policies and populations: Implications for children, families, and policymakers. *Children and Youth Services Review*, 57(C), 30-39. doi: 10.1016/j.childyouth.2015.07.008
- Miller, A. (2013). Potential Dangers of Unlicensed Daycares. *Canadian Medical Association Journal*, 185(18), 1566. doi: 10.1503/cmaj.109463
- Moloney, M. (2016). Childcare regulations: Regulatory enforcement in Ireland. What happens when the inspector calls? *Journal of Early Childhood Research*, 14(1), 84-97. doi: 10.1177/1476718X14536717
- National Center on Child Care Quality Improvement. (2013). Research Brief #1: Trends in Child Care Center Licensing Requirements and Policies for 2011. Retrieved from https://childcareta.acf.hhs.gov/sites/default/files/public/065_1304_center_licensing_trends_brief_final.pdf
- Rosenthal, M. S., Jeon, S., & Crowley, A. A. (2016). Health and Safety in Family Day Care Homes: Association Between Regulatory Non-compliance and Lower Median Income. *Maternal & Child Health Journal*, 20(5), 984-992. doi: 10.1007/s10995-015-1883-y
- Strain, P. S., Barton, E. E., & Dunlap, G. (2012). Lessons learned about the utility of social validity. *Education & Treatment of Children*, 35(2), 183-200. doi: 10.1353/etc.2012.0007
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. Jossey-Bass, an imprint of Wiley.
- Washington State Department of Children, Youth & Families (formerly Department of Early Learning). (n.d.). Become a Licensed Child Care & Early Learning Provider. Retrieved from <https://www.dcyf.wa.gov/services/early-learning-providers/licensed-provider> (formerly <http://www.del.wa.gov/care/about/Default.aspx>)

- [29] Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203-214. doi: 10.1901/jaba.1978.11-203