Strengthening Texas Rising Star Implementation Study

INTERIM UPDATE FROM THE CHILDREN'S LEARNING INSTITUTE AT UTHEALTH

SEPTEMBER 26, 2019

Agenda

- Interim Executive Summary Review (1.5 hours)
- Parking lot discussion

Final Report Resources

- Executive Summary
- Final Report
  - May include additional analyses
- Appendix
  - Item revision table

Study Updates to TWC

- Monthly Conference Calls
- 3 in-person meetings
  - July 2018 (preliminary data)
  - May 6, 2019
  - September 4, 2019

Study Aims

▶ Aim 1: To examine the reliability of the TRS assessment. This was the primary aim of data collection and is intended to provide key evidence to support removal or revision of measures.

▶ Aim 2: To examine indicators of external validity of the TRS assessment across categories and with other measures of quality and outcomes.

▶ Aim 3: To examine qualitative aspects of implementing TRS assessment training and data collection to determine the impacts of scoring rules and assessment procedures on reliability and system efficiency.

Study Sample

- **Final study sample: 128**

- **Classrooms assessed: 864 total**

| Age Groups	Low	Medium	High	Total
Infant (0-17 months)	58	72	59	**189**
Toddler (18-35 months)	69	96	82	**247**
Pre-K (3-5 years)	68	113	99	**280**
School Age (5-12 years)	44	62	42	**148**
Total	**239**	**343**	**282**	**864**

Study Limitations

- Recruited licensed center-based child care facilities that served all ages only

- Findings are not necessarily representative of centers that serve a limited age population (e.g., school-age only) or home-based child care providers
  - Recommended to separately study reliability and validity in home-based child care

- Initial exploration of validity was limited given our primary focus on reliability and development of certification procedures
  - Recommend the collection of more extensive and diverse validity evidence after field reliability is established

Key Definitions for Analysis & Recommendations

Indicators of Reliability

▶ **Normality of score distribution**: A method of examining item functioning. Item scores can be normally distributed or skewed (i.e., scores concentrated at the low or high ends). Highly skewed items fail to differentiate quality among providers assessed, which contributes little information to the assessment system and results in missed opportunities to capture rich data.

▶ **Internal consistency**: A measure of instrument reliability that determines if items within the same category and subcategories measure the same concepts. Internal consistency values greater than .6 are considered acceptable for research purposes. Values above .90 are considered excellent and are the desired level.

▶ **Inter-rater agreement**: A measure of rater reliability that indicates the extent to which two people scoring side-by-side are able to reach the same rating.
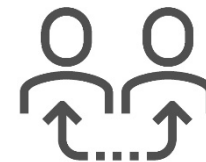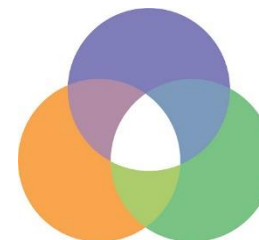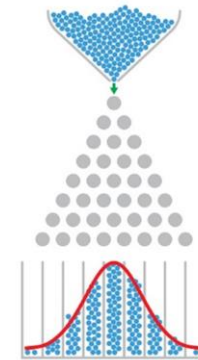
Key Definitions for Analysis & Recommendations continued

▶ **Generalizability coefficient:** A measure of rater reliability that indicates the extent to which a team of raters draw similar conclusions, accounting for differences across the raters and sites assessed.

Results & Category-Level Recommendations

By category:

- Category description
- Study highlights
- Selected results that inform recommendations
- Recommendations

Category 1: Overview

Subcategory	Number of Met/ Not Met Items	Number of Points-Based Items
Director Qualifications	2	3
Caregiver Qualifications	6	2

Items relating to:

▶ Caregiver education, experience, and staff training

▶ Director education, experience, and staff training

Category 1: Data (structural measures)

- No center met all category 1 requirements for a 2-star rating

TRS Evaluation study: Structural Rating Percentages
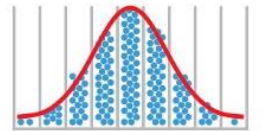Category1
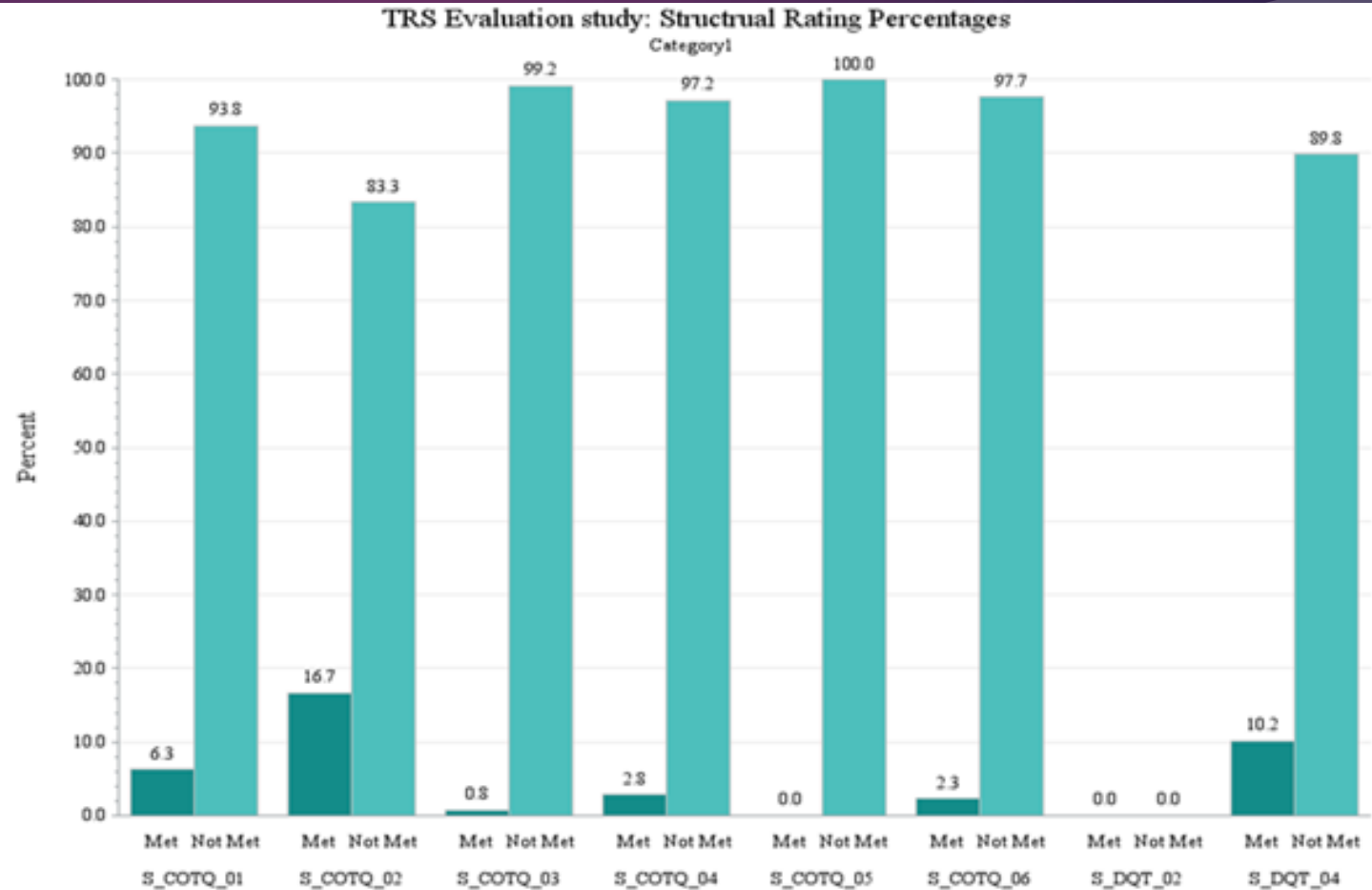
Category 1: Data (excluded n/a)

- Data for a high number of facilities was excluded (i.e., scored "not applicable")
  - S_COTQ_02 volunteer and substitute caregiver orientation, 86%
  - S-COTQ-04 full-time caregiver staff training-school age, 45%* (N/A allowed if caregiver employed for less than 90 days)
  - S-COTQ-05 part-time caregiver staff training- school age, 61%* (N/A allowed if caregiver employed for less than 90 days)
  - S_DQT_02 TRS director certification course, 100%

Category 1: Highlights

▶ No center met all category 1 requirements for a 2-star rating

▶ Data for a high number of facilities was excluded (i.e., scored "not applicable")

▶ Several item-level indicators are difficult to consistently capture

▶ Category 1 is time-intensive for assessors to score
  ▶ In study: 30-40 minutes per caregiver/director, up to 90 minutes
  ▶ Using TECPDS reports: 10-15 minutes per caregiver/director

▶ Study team developed worksheets that better facilitate scoring

▶ Required key elements were more easily scored using TECPDS individual profile reports

Category 1: Recommendations

- Recommend to revise or remove item-level indicators that:
  - **have a high rate of N/A scores**, unless the indicator is strongly supported by theory and/or evidence;
  - **do not differentiate provider quality** (i.e., highly skewed scores), which will lessen the burden on providers and assessors and reduce the amount of time required to complete an assessment; and
  - **are inconsistently captured and available for review**. Conversely, setting new field expectations and norms for including this information in routine document issuing and management practices.

Category 2: Overview

Number of Items by Age Group and Subcategory

Subcategory	Infants	Toddlers	Preschool	School-age
Staff Ratios and Group Size	1	1	1	1
Language Facilitation and Support	10	10	10	10
Play-Based Interactions and Guidance	3	3	3	3
Support for Children's Regulation	0	7	7	7
Warm and Responsive Style	6	6	6	6

Items relating to:

▶ Group size

▶ Caregiver-child ratio

▶ Quality of interactions between caregivers and children
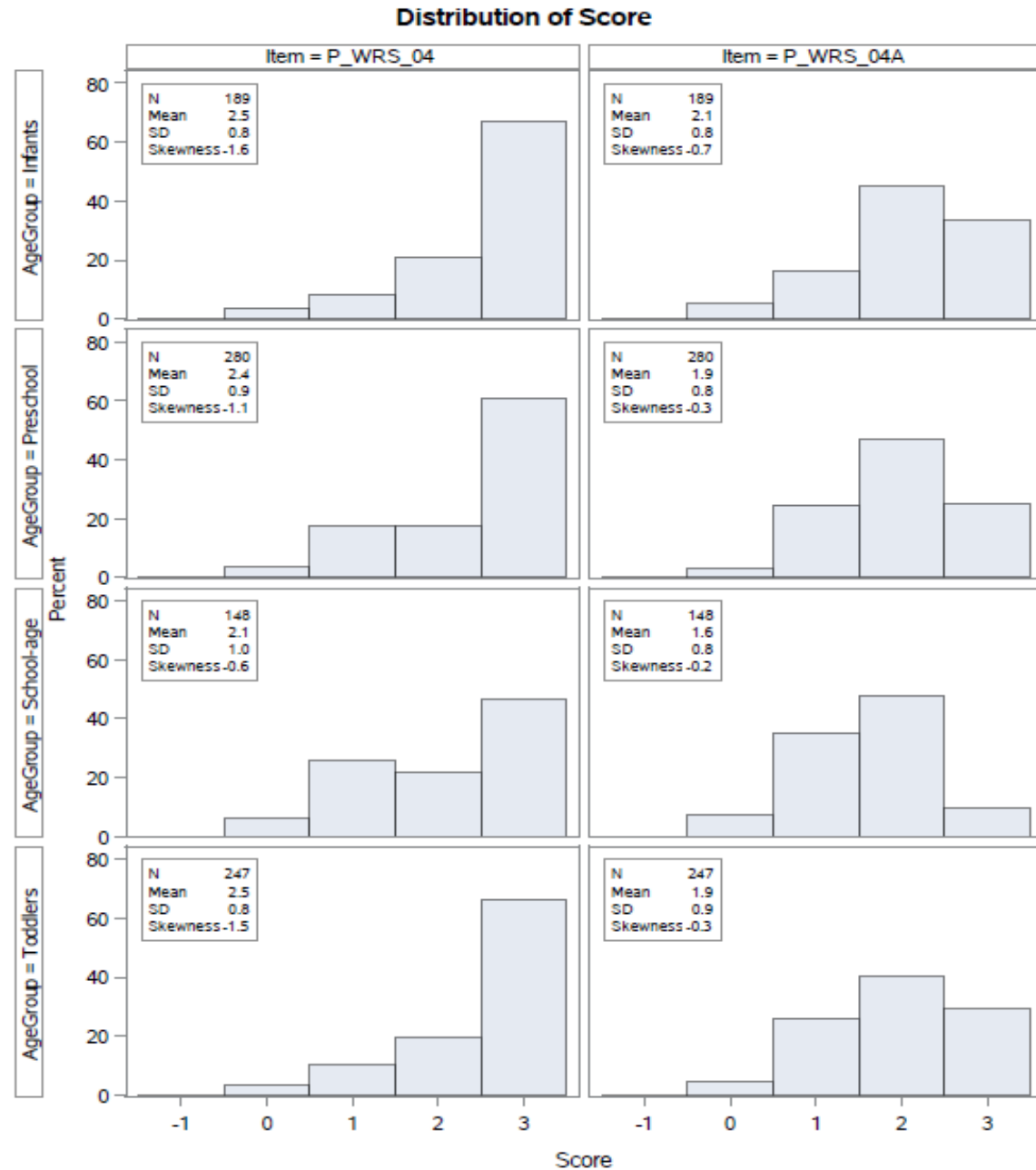
Category 2: Data (Attendance)

	Score of 0	Score of 1	Score of 2	Score of 3
Enrollment Information Review	23%	19%	22%	35%
Present during Assessment	9%	13%	22%	57%

Category 2: Data (Ratio)

Pearson Correlation

Category 2 CTR	Group Ratio Enrolled	Group Ratio Present
Infants	-0.17613	-0.20039
Toddlers	-0.09843	-0.19135
Preschool	-0.16847	-0.16743
School-age	-0.13778	-0.12784

Category 2: Data

Category 2: Data (Internal Consistency)

Age Group	Traditional	Alternate
Infants	0.90	0.93
Toddlers	0.91	0.93
Preschool	0.91	0.93
School-age	0.90	0.92

Cronbach Alpha > 0.70 highlighted

Category 2: Highlights & Recommendations

▶ With rigorous training, assessment team reached reliability

▶ Study examined differences in scores for group size/ratio using enrollment data vs. staff and children present and recommend adjusting scoring criteria for group size

▶ Identified alternate scoring that results in greater reliability for category 2 items currently scored by frequency counts of behaviors

▶ Internal consistency for category 2 for all items using both current and alternate scoring methods is in the excellent range (.90 and above) for all ages.

Category 3: Overview

Number of Items by Age Group and Subcategory

Subcategory	Infants	Toddlers	Preschool	School-age
Instructional Formats and Approaches to Learning	5	5	5	5
Lesson Plans & Curriculum	4	4	10	1
Planning for Special Needs & Respecting Diversity	3	3	3	3

Items relating to:

▶ Curriculum, including lesson plans and instructional formats

▶ Planning for special needs

▶ Considerations for children from bilingual and culturally diverse backgrounds

Category 3: Data

- ▶ Curriculum and Lesson Plan scores for nearly every item, across age groups, show floor effects (i.e. most classes receive a score of 0).

- ▶ Scores within Instructional Formats and Approaches to Learning were more normally distributed.

- ▶ Scores for P-SNRD show floor effects and a high percentage of caregivers were excluded from rating with a "not applicable" score.

Category 3: Data continued

- Internal consistency:
  - Infants (.66 and .69, respectively) borderline acceptable and
  - toddlers (.60 for both scoring methods) borderline acceptable
  - Internal consistency for preschool items reaches the good range for both current and alternate (.85 and .81).
  - School-age internal consistency is unacceptable for both scoring approaches (.51 and .47).
- IFAL show stronger item characteristics and shows significant moderate to large correlations with Category 2 (r= .42 to .56, p<.01).

Category 3: Highlights

- ▶ Category 3 is not functioning well in internal consistency and distribution of scores.

- ▶ Internal consistency for category 3 only reaches the good range for preschool items.

- ▶ As currently written, lesson planning is not providing a strong measure of curriculum

Category 3: Recommendations

▶ Recommend removal of lesson planning items:

    o Difficult to achieve initial reliability

    o Most time-intensive items to score (average 30-45 minutes per classroom)

    o Lack of evidence to support this approach to measuring curriculum

▶ Recommend removal of planning for special needs and respecting diversity items as currently measured

▶ Recommend moving IFAL items to category 2

    o Correlations suggest IFAL and category 2 may be appropriately scored together
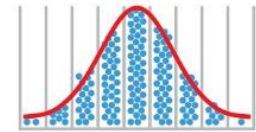
Category 4: Overview

Number of Points-Based Items by Age Group and Structural Items by Facility within each Subcategory
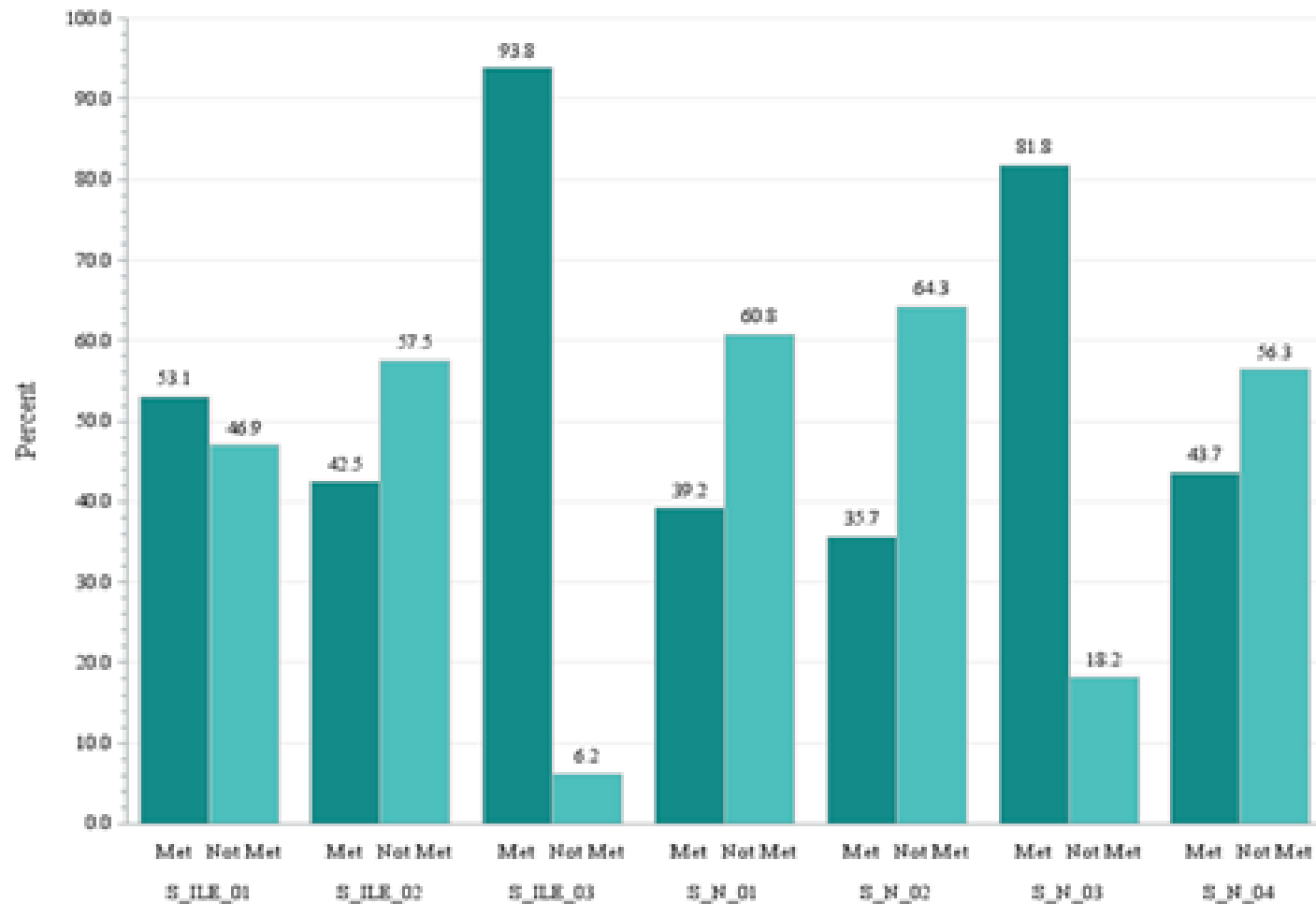
Subcategory	Infants	Toddlers	Preschool	School-age	Facility Met/Not-Met
Indoor Learning Environment	7	7	7	8	0
Nutrition	3	3	4	3	4
Outdoor Learning Environment	5	4	4	4	0

Items relating to:

▶ Nutrition

▶ Indoor learning environment

▶ Outdoor learning environment

Category 4: Data (structural)

TRS Evaluation study: Structural Rating Percentages
Category4

Category 4: Data

- With few exceptions, item level distributions for points-based measures within Category 4 were acceptable. Some notable concerns included:

- Item P-N-01 considers to 6 specific mealtime practices and scored at a 3 for more than 83% of providers

- Item P-N-03 was often excluded (38%) because the majority of children in the infant classrooms were receiving solid foods.

- Item P-N-04 was often excluded (37%) because all children in the observed infant classroom were above 12 months of age.

- P-OLE-01 which considers the extent to which the outdoor environment activities are linked to indoor learning was score as 0 for 80% providers

Category 4: Data continued

- Internal consistency:

  - Infant items is borderline acceptable (.60).

  - Toddler, preschool, and school-age items show internal consistency in the acceptable range (.79 to .80).

- Internal consistency after item removal was acceptable (.80):

- P-N-03 and P-N-04 from infant items

- P-N-01, P-N-02, and OLE-01 from school age items

Category 4: Highlights

- Several items showed limited variation in score, indicating they do not differentiate quality

- Nutrition contains too few items to be able to fully assess reliability, and several items show limited variation

- Indoor learning environment items (across all ages) shows acceptable reliability

- Outdoor learning environment items shows acceptable reliability for all ages except infants

- No notable differences in internal consistency for the current and alternate scoring methods
  - Infant items: borderline acceptable
  - Toddler, preschool, and school-age items: acceptable range
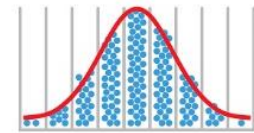
Category 4: Recommendations

▶ Recommend removing category 4 items with limited variation in score

   ▶ Lessen the burden on providers and assessors

   ▶ Reduce assessment time

▶ Recommend removal of low performing nutrition items

   ▶ Removing these items resulted in improved category 4 reliability

   ▶ Nutrition practices recommended for the continuous quality improvement framework (CQI; recommendation 6)

Category 5: Overview

Subcategory	Number of Met / Not Met Items	Number of Points-Based Items
Parent Education	2	2
Parent Involvement	3	3

Items relating to:

- ▶ Education and involvement of parents and other family members in the program

Category 5: Data

TRS Evaluation study: Structrual Rating Percentages
Category5
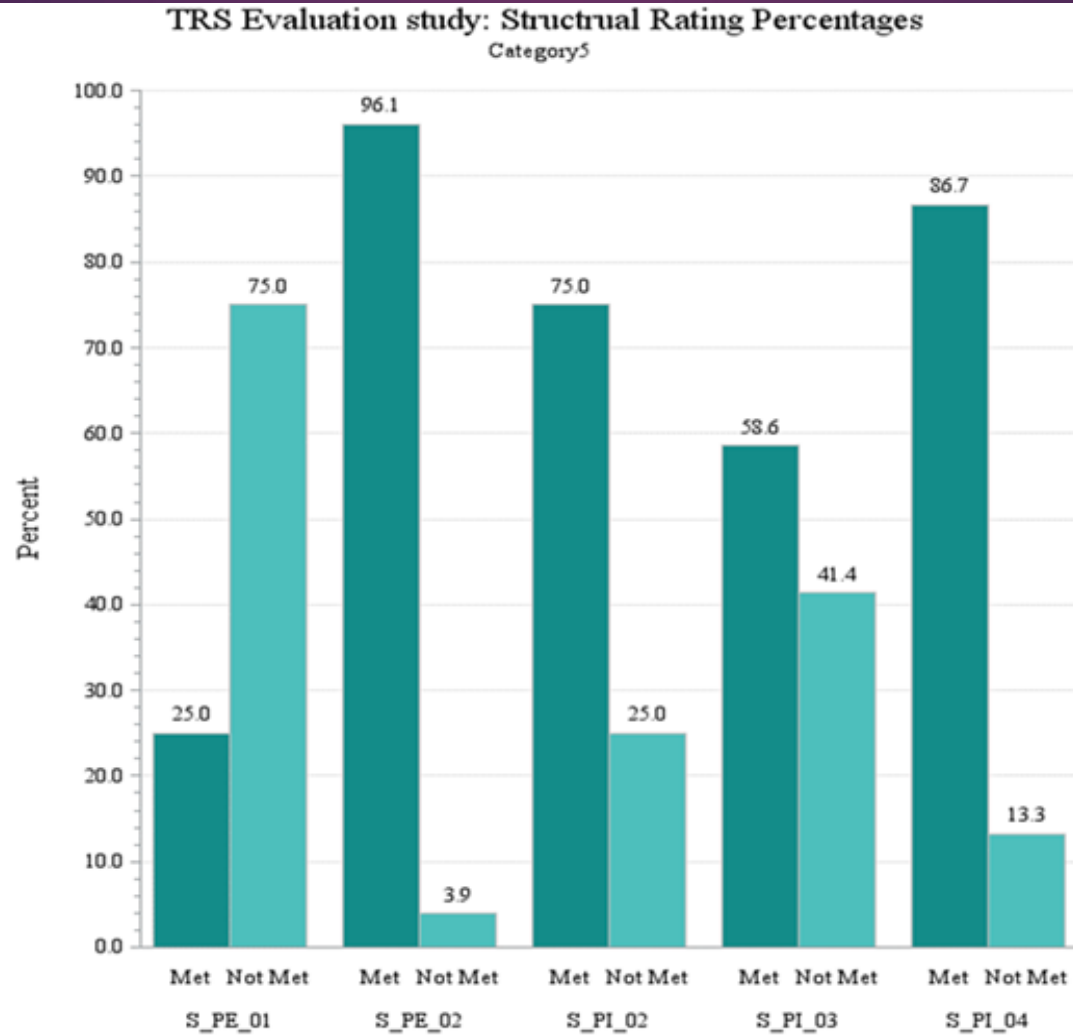
Category 5: Highlights

- Several items do not involve objective review of evidence, instead rely heavily on self-report

- Few items showed limited variation in score

  - Example: 96% of providers met S-PE-02 (school-parent communication system)

- Category includes a small number of items, and only acceptable reliability was established

- Internal consistency is in the borderline acceptable range (.70) and item removal not examined for this category

  - Items are normally distributed

  - All items correlate moderately with the total score
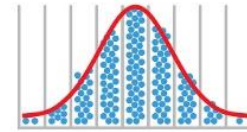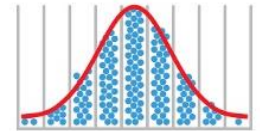
Category 5: Recommendations

- Recommend removal of items with limited variation in score
  - Recommend removal of S-PE-02
- Recommend adjusting weight of category 4 within the overall star rating calculation when further validity data becomes available

Cross-Category Findings and Recommendations

- Adjustments to categories based on item-level screening procedures

- Completed factor analysis to confirm underlying constructs within the TRS assessment

- Compared generalizability coefficients, internal consistency, distribution of star ratings, and stability of ratings over time

  - Using the current and recommended structures

- **Convergence in evidence** across multiple analytical approaches improves our confidence that recommended changes will improve performance of the TRS assessment

Key Definitions for Cross Category Findings & Recommendations

- **Factor analysis**: A statistical method used to explore or confirm the number of underlying constructs (i.e., concepts measured by the TRS assessment) and examine the extent to which the items are designed to measure the same construct
  - This analysis **increases confidence** that items within categories measure the constructs the TRS program intends to measure
- **Generalizability coefficient:** A measure of rater reliability that indicates the extent to which a team of raters draw similar conclusions, accounting for differences across the raters and sites assessed.

Recommended Structure: Confirmatory Factor Analysis

- Category 3 Lesson planning and curriculum items were measuring one construct in the preschool age group only

- Category 2 results show Results indicated a one-factor structure fitted data well in four age groups
  - IFAL included in the factor structure for category 2

- Category 4 results confirmed three separate dimensions exist within this category
  - Category 4 does not measure a single construct

- Category 5, the results showed an one-factor structure fitted data well
  - Suggesting final items of this category were measuring one construct

Overall Internal Consistency for Points-Based Items

- Internal consistency was strong using the current structure
- Analyzed recommended structure with Cronbach's alpha for points-based items and found small improvements across all age groups
  - Infant items with recommended structure: internal consistency was improved from the "good" to "excellent" range
  - Result: "excellent" internal consistency for all age groups
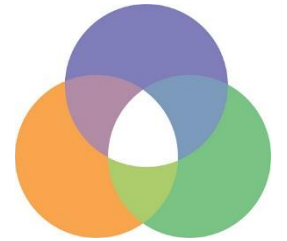
Inter-Rater Reliability

- Generalizability coefficient (G-coefficient) was estimated overall for all points-based, classroom-level items in categories 2, 3, and 4 for the current and alternate scoring methods
  - Rater-level reliability under current scoring ranging "marginally acceptable" range to "acceptable" range
  - G-coefficients were slightly higher for the alternate items (.71 to .89)
  - Reliability for study rating team (10 members):
    - Six assessors in the "acceptable" range
    - Three in the "relatively acceptable" range
    - One rater failed to maintain reliability and was reassigned
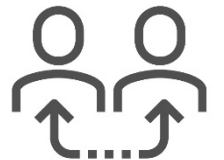
Inter-Rater Reliability continued

- Examined G-coefficients under the recommended measure structure (categories 2, 3, and 4) for all points-based, classroom-level items
  - Generalizability coefficients were slightly higher with the new structure, with 8 of the 9 raters in the "good" to "excellent" range (one rater remained in the "acceptable" range).
  - Supports using the recommended measure structure to improve accuracy and reliability of field ratings
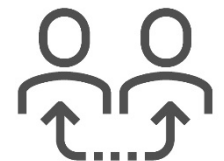
Distribution of Star Ratings

- No study sample providers met all of the requirements for 2-star certification

- Percentage of providers with met/not met ratings within categories:

  - Category 1: no providers met all met/not met items

  - Category 4: only three providers (2%) met all met/not met items

  - Category 5: , 23 providers (18%) met all met/not met items

- Many items that require providing documentation or self-reporting could meet these requirements using standardized templates and sample documents
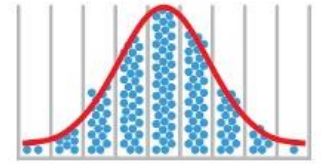
Distribution of Per Category Star Ratings

- Excluded met/not met items to examine variation in star ratings based on points-based items
  - See category variation table at right
- We also examined the distribution of star ratings under the recommended structure (i.e., excluding items recommended for removal), and found no changes in overall star rating and very few changes within category scores.

Number of Providers Per Category Star Rating (excluding met/not met indicators)

Category	2-Star	3-Star	4-Star
1	115	12	1
2	114	14	0
3	128	0	0
4	110	18	0
5	79	28	21

Initial Exploration of External Validity

▶ Primary scope of the study to examine for and support reliability

▶ Where possible:

- ○ Examined for relations across categories
- ○ Among TRS items
- ○ External sources that provide initial evidence that TRS scores correlate with other aspects of quality

Q1: Are star ratings stable across brief periods of time?

▶ Stability of ratings measured changes in category and overall star ratings in between repeated assessments of the same providers.

▶ The study examined 40 facilities and 269 classrooms as part of the stability rating sample.

Study Sample

# of facilities	# of classrooms	# of assessment visits	Period of time
24 facilities	164	2	2.5 weeks after assessment 1
16 facilities	105	3	8.2 weeks after assessment 2

Q1: Are star ratings stable across brief periods of time?

► The results showed:
  - Overall star ratings were stable across time.
  - At the category levels, star ratings were typically stable.

Category Level		
Categories	Assessment Visit 2 (n=11)	Assessment Visit 3 (n=1)
1	No change	No change
2	3 facilities	No change
3	No change	No change
4	6 facilities	1 facility
5	2 facilities	No change

Q1: Are star ratings stable across brief periods of time?

▶ Stability was more of a concern at the classroom level

▶ Changes in caregiver were frequent in our sample, even over relatively brief periods of time.

- Sixty-six percent of classrooms had a stable **lead** caregiver across three assessments.

- Fifty-nine percent of classrooms had stable caregiving staff (including both lead and co-caregivers) between assessments 1 and 2.

- Thirty-eight percent retained the same classroom makeup across three assessments.

It is worth noting that many children in the centers in the study sample are not experiencing continuity of care, which may make it difficult for children to build relationships with individual caregivers.

Q1: Are star ratings stable across brief periods of time?

▶ In the subsample of classrooms (n=40) that retained the same classroom makeup (i.e., all caregivers the same across time), there were small but significant decreases in scores over time (between assessment 1 and 2 and 2 and 3).

▶ We re-examined stability across time for all 269 classrooms using the recommended structure and found that the differences for caregiver-child interactions for observations 1 and 2 were still significant, but the differences between observations 2 and 3 (for 105 classrooms) were no longer significant. This suggests that scores are more stable under the recommended structure.

Q2: Is there evidence that star ratings and classroom quality vary by socioeconomic status?

Met/Not Met

▶ Only a few items with identifiable SES differences

▶ Most providers, regardless of SES, scored Not Met on most indicators

Points-based

▶ *Slight* trend toward higher scores within higher SES providers

▶ In the highest rated SES group, providers on average would not meet the threshold for a 3- or 4-star rating at the category level

Q3: Is accreditation related to TRS scores?

▶ Study sample included 18 accredited providers, all of which received a full site assessment.

▶ None scored at the 4-star level on points-based items

▶ Scores for accredited providers were slightly higher than non-accredited providers for categories 2, 4, and 5, but these differences were not substantial enough to change overall star ratings.

▶ Based on this sample of providers, we did not find evidence to support automatic 4-star ratings for nationally accredited programs.

Q4: Do directors with higher levels of education, training, and experience have higher scores on TRS facility scores?

- We examined for correlations between all category 1 director-focused items and TRS classroom items
  - No consistent patterns
- We also looked at the extent to which individual indicators (e.g., years of experience, business management training hours) relate to classroom and facility points at the category level
  - We found multiple small to moderate significant correlations with facility-focused categories
- This suggests information is lost with the current item structure, which may limit it's ability to predict outcomes

Q5: Do caregivers with higher levels of education, training, and experience have higher scores in caregiving behaviors?

We examined correlations between all category 1 caregiver-focused items and TRS classroom items and found a fairly consistent pattern of correlations that suggests:

▶ Providers with more qualified staff (measured by P-CQT-01) have small to moderate correlations with higher scores for category 2 and category 4, and higher category 4 star ratings.

▶ Caregiver staff training topic alignment (measured by P-CQT-03) is moderately related to category 3 scores.

Q6: Do lower caregiver-child ratios relate to higher TRS scores?

▶ Low caregiver-child ratios are widely considered to be an important structural feature of quality programs, that allows caregivers to better supervise children and engage in more positive interactions

▶ In the study sample, better scores for TRS group/ratio shows significant small correlations with category 2 and 4 scores

▶ We also looked to see if ratio relations were stronger for certain age groups within each category and found that correlations were small across all age groups

Q7: Do TRS scores for caregiving behavior (e.g., category 2) relate to another established measure of caregiving quality (convergent validity)?

▶ We examined for evidence of convergent validity by comparing TRS scores for caregiver-child interactions with scores from another established measure of caregiver interaction quality, the Arnett Caregiver Interaction Scale

▶ Multiple high significant correlations were found with category 2 scores and Instructional Formats and Approaches to Learning (in category 3) than with non-behavioral items

▶ These data provide initial evidence that the behavioral observation items within the TRS assessment relate well to other measures in routine use

Q7: Are TRS scores for caregiving behavior (e.g., category 2) sensitive to changes in caregiving quality in the context of intervention (e.g., before and after CQI)?

- Toddler Pilot Sample
  - 40 teachers in Dallas and Houston (20=target, 20=control)
  - 8% 4-year degree, 73% high school only; 18% other
  - 8% Caucasian, 55% African-American, 25% Hispanic/Latino, 18% other
- Average of 6 children per classroom, ages 24-36 months
- Total number of children 241 (115 control and 126 intervention)
  - 52% boys and 48% girls
  - 55% Black, 23% Latino, and 21% other
  - Family income 34% $30,000 or below, 23% between $30,001 and $60,000, and 29% higher than $60,000

Toddler Pilot: Intervention Delivery

- Target teachers received (over approximately 6 months):
  - language, literacy, and social-emotional online courses;
  - a face-to-face kickoff training;
  - classroom kits that support skill-building; and
  - four hours of individualized remote coaching per month.
- To promote self-reflection, teachers were assigned activities to film and upload for coaching feedback.

Toddler Pilot: Impacts on Category 2

	b	(SE)	*p*-value	Effect Size
Category 2 Total	**9.81**	**(4.99)**	**0.049**	**0.58**
Language facilitation/support	**6.01**	**(2.08)**	**0.004**	**0.80**
Play-based interaction/guidance	0.93	(0.81)	0.251	0.32
Support for children regulation	1.51	(1.22)	0.213	0.40
Warm/responsive style	2.41	(1.85)	0.192	0.45

Key Recommendations

- Goals for QRIS:
  - **A market-based system for improving quality**
  - **Workforce professionalization**
  - **Support for child care providers**

Item Recommendations	Reliability	Validity	Training	Implementation
Recommendation 1: Removing or adjusting low-performing items to improve instrument functioning	X		X	X
Recommendation 2: Adjusting the relative weight of categories to be more in line with measure reliability and to more accurately reflect the influence of evidence-based practice on children's outcomes	X	X		
Recommendation 3: Revising procedures for automatic certification of nationally accredited providers		X		
Recommendation 4: Employing a rigorous training and reliability monitoring process to ensure accurate star rating across the state	X	X	X	X
Recommendation 5: Standardizing application and scoring routines to improve program efficiency and accuracy of star assignment	X		X	X
Recommendation 6: Establishing a quality improvement framework that uses a developmental approach to ensure providers receive technical assistance and professional development in alignment with their current star ratings	X	X		X
Recommendation 7: Continuing exploration of external validity.		X		

Recommendation 1:
Removing or adjusting low-performing items to improve instrument functioning.

- Recommending retention of approximately 71% of the current items
  - Recommend revising scoring criteria and/or updating the technical scoring manual (TSM) for approximately 35 items
  - Study successfully tested the recommended alternate scoring
  - Recommend minor TSM updates only for an additional 10 items
- For the remaining 29% of items, recommend item removal based on analysis results, implementation concerns, or both
  - Study uncovered implementation concerns
  - Item removal recommendations are primarily related to lesson planning, nutrition, indoor learning environments, and parent education

Recommendation 2:
Adjusting the relative weight of categories to be more in line with measure reliability and to more accurately reflect the influence of evidence-based practice on children's outcomes.

- TRS system currently has five categories that receive equal weight in star rating calculation
  - Current scoring approach signals equal importance for all categories of quality
- Aspects of care linked to children's experience and outcomes should feature prominently in the quality rating
  - Measurement of child outcomes is beyond study's scope, but constructs aligned with some TRS categories are more closely related to children's experiences and outcomes
  - Research shows high quality learning experiences within warm and responsive relationships with adults advances outcomes for at-risk children
    - These *process* features of care are consistently found to be stronger predictors of student outcomes than *structural* features of care, such as director qualifications
  - Caregiver-child interactions (category 2) and instructional formats and approaches to learning (subcategory of category 3) are highly aligned with better child outcomes
    - Items that performed well during the study should be the most heavily weighted in the rating

Recommendation 2:
Adjusting the relative weight of categories to be more in line with measure reliability and to more accurately reflect the influence of evidence-based practice on children's outcomes.

- Example approaches for adjusting the relative weight:
  - Assign differential weights to each category to align with the evidence base (e.g., category 2 would receive more weight than category 5)
  - Calculate average scores across all items in the recommended assessment structure, which would place more weight on caregiver-child interactions because of the higher number of items measuring this construct.
- Establish statewide reliability using the recommended structure, followed by a validity study (including predictive analysis) that captures key outcomes aligned with TRS goals

Recommendation 3:
Revising procedures for automatic certification of nationally accredited providers.

- Within the study, no accredited providers were scored at a 4-star level
- Data suggests changes for participating nationally accredited providers:
  - Discontinue automatic 4-star ratings
  - Conduct full site assessment prior to certification
- Aid targeting efforts in continuous improvement plans

Recommendation 4:
Implementing a rigorous training and reliability monitoring process to ensure accurate star rating across the state.

- Inter-rater reliability has significant implications for the fairness of quality ratings attributed to providers and the accuracy of ratings communicated to families
- Study assessors were able to reach "acceptable" inter-rater reliability after a rigorous training process
- Adopt a rigorous training process using research-supported standards and procedures to ensure accurate ratings across the state
- Consider adopting an accountable central body that certifies reliability and conducts routine reliability monitoring
  - Centralizing reliability certification and monitoring of staff across the state ensures reliable assessment of fair and accurate ratings
- TRS assessors should be required to maintain a monthly minimum of classroom observations
  - Maintaining reliability requires frequent and consistent use
- Establish monitoring procedures to capture assessor "drift" and prompt re-training efforts
- Require specific notetaking and documentation procedures to bring clarity to the ratings process, strengthen the accuracy of ratings, and provide evidence for specific scores in communications with providers

Recommendation 5:
Standardizing application and scoring routines to improve program efficiency and accuracy of star assignment.

- Require specific notetaking and documentation procedures to strengthen ratings and score communication

- Establish new program-level portfolio submission requirements to enable review and incorporation into quality improvement plans

- Require assessors to utilize TECPDS reports to facilitate scoring of director and caregiver qualifications

- Integrate the TRS Interest Form, TRS Application, and TRS Provider Certification Screening Form with the TRS Assessment to streamline information collection and scoring

- Integrate TECPDS with the TRS Online Assessment Tool for automated scoring of director and caregiver qualifications

Recommendation 6:
Establish a continuous quality improvement (CQI) framework that uses a developmental approach to ensure providers receive technical assistance and professional development in alignment with their current star ratings.

- ▶ CQI approach can be used to target technical assistance to *lift quality and increase participation* in the program, leveraging TRS mentoring staff:
  - ▶ Before certification, if providers are not able to meet TRS standards
  - ▶ Move existing certified providers to *progressively higher* levels of quality that fully meet TRS expectations
  - ▶ Technical assistance in specific areas to *maintain* quality for providers already delivering high quality services
- ▶ Coordinated CQI framework can include a combination of self-study materials, professional learning communities (PLCs), and individualized coaching for providers at all levels of quality and stages of TRS implementation
- ▶ Some areas of TRS standards identified for removal based on data and implementation concerns can be meaningfully incorporated in a CQI approach
  - ▶ Items related to lesson planning, planning for special needs and respecting diversity, and nutrition were found to function poorly, cumbersome for assessors to rate, or did not differentiate quality
  - ▶ Items may be more appropriately measured and addressed through the use of school leader and staff interview protocols
    - ▶ Establish qualitative scores that can be used by mentors to support providers in implementing CQI plans
- ▶ Require mentors to monitor quality improvement fidelity metrics
- ▶ Adjust CQI plans based on most recent assessment data (e.g., annual monitoring)

Recommendation 7:
Continuing exploration of external validity.

- Study focused on strengthening the reliability of TRS ratings to:
  - Ensure reimbursement rates are accurately allocated
  - Technical assistance is appropriately targeted to needs
- Study found some initial evidence of validity (e.g., strong correlations between TRS caregiver-child interactions and validated measures of caregiving quality)
  - Once field reliability is established using the recommended structure, additional research is recommended to further examine:
    - Long-term rating stability
    - Ability of the CQI approach to increase TRS participation and advance providers to increasing levels of quality
    - Evidence that participation predicts outcomes of interest/TRS goals

Next Steps

▶ For your consideration: how will the findings help you make recommendations to TWC?

▶ Send questions to: TRS4YearReview@twc.state.tx.us

Thank You!

LaShonda Brown-Hollins

LaShonda.Brown@uth.tmc.edu

Jennifer Lindley

Jennifer.M.Lindley@uth.tmc.edu