Validation of the CONNECT Survey Co-Designed with Parents to Strengthen Parent/Caregiver-Clinician Relationships in Pediatric Primary Care

June 30, 2025

Reach Out & Read®

Institute for Child Success

Executive Summary

The parent/caregiver experience of pediatric primary care is a critical factor that influences child health outcomes and development. Existing patient experience assessment tools are often designed for adult healthcare settings and may not capture the relational elements that matter most to families. To address this gap, Reach Out and Read, in collaboration with the Institute for Child Success, developed the CONNECT Survey – a concise 13-item measure co-created with parents to assess the relational aspect in a well-child visit. Designed through an iterative process involving empathy interviews and caregiver-led Delphi panels, the survey reflects family priorities using accessible language while remaining practical for routine clinical use.

Between July and September 2024, a total of 9,825 caregivers from 131 Reach Out and Read clinics across five Mid-Atlantic states completed the CONNECT survey. An additional 60 caregivers in Florida completed the survey twice across two weeks to check test-retest reliability. The sample reflected the diversity of Medicaid-serving pediatric practices: 58% publicly insured, 44% White, 19% Black, 25% Hispanic, and representation from urban, suburban, and rural communities.

Key Findings

Reliability: The survey demonstrated exceptionally high internal consistency (Cronbach's $\alpha = 0.993$; McDonald's $\omega = 0.993$), with stable performance across racial, insurance, and child-age subgroups—confirming that all items reliably measure a single concept: relational quality.

Construct Validity: Exploratory factor analysis identified a single dominant factor accounting for 99% of score variance, indicating strong unidimensionality.

Convergent Validity: Scores correlated significantly with established patient experience items from CAHPS and Press Ganey ($p \approx 0.23$–0.30), supporting the survey's alignment with related constructs.

Concurrent Validity: Caregivers who regularly saw the same care provider and brought the child to visits reported higher CONNECT scores, supporting the survey's ability to capture meaningful relational dynamics in pediatric care.

Test-Retest Reliability: Among caregivers who completed the survey twice, item agreement ranged from 90% to 96%, with Gwet's AC1 values of 0.87–0.95—indicating strong temporal stability.

Equity Insights: While overall scores were high, White and privately insured families consistently reported the most positive experiences. In contrast, Black, multiracial, publicly insured, and preschool-age families reported modestly lower ratings, highlighting important opportunities for equity-focused improvement.

CONNECT is brief, easy to administer, and designed for real-world use—whether integrated into clinic workflows or distributed via patient portals. Results can be reviewed monthly to detect trends, guide improvement cycles, and support clinician development. In an evolving healthcare landscape that increasingly rewards patient experience, CONNECT offers a reliable, pediatric-specific, and equity-sensitive metric of relational care. Future work will explore refinements to reduce ceiling effects, streamline overlapping items, and expand testing across more diverse pediatric settings.

Table of Contents

Introduction

Co-created *by* parents *for* parents, the **CONNECT** *(Clear Communication, Open Dialogue, Needs Acknowledged, Nurture Connection, Empathy and Respect, Collaborative Decisions, Trust)* survey was developed by a partnership of Reach Out and Read and the Institute for Child Success (ICS) to gain insights into the parent/caregiver-clinician relationship during well-child visits.

The Importance of Relationships in Pediatric Primary Care

Positive childhood experiences (PCEs) are essential to children's healthy development, offering protection against the harmful effects of adversity, such as poverty and racism. Recent research highlights that even children in difficult circumstances can flourish emotionally and academically when supported by strong relationships within families and with healthcare providers. This has led to the concept of "Early Relational Health" (ERH)—the quality of early relationships children experience—as both a measurable indicator of child well-being and a target for positive interventions (Garner & Yogman, 2021; Bethell et al., 2022; Willis & Eddy, 2022; COHC & IPFCC, 2012; Petts & Shahidullah, 2020).

Reaching nearly all families through routine well-child visits (up to 14 visits from birth to age five), pediatric primary care is uniquely positioned to foster these beneficial relationships and experiences. Additionally, pediatric care is among the most equitable healthcare settings, regularly engaging diverse families. Interventions delivered in primary care that focus on family strengths and cultural respect have increasingly gained recognition.

Recent research demonstrated that parents and caregivers welcome guidance on ERH during routine pediatric visits in the context of a good relationship with their pediatric care provider (Cordoba et al., 2024; Center on the Study of Social Policy, 2020). As a construct that fits within parent experience, this finding is in agreement with the extensive literature showing that patient experience is crucial for high-quality medical care. Since 2008, enhancing patient experience has been recognized as one of the essential goals ("Triple Aim") needed to improve the US healthcare system, alongside improving overall population health and reducing healthcare costs (Berwick, Nolan, & Whittington, 2008). Specifically for pediatric care (Burton & Navasaria, 2019), the American Academy of Pediatrics (AAP) has emphasized that family experience significantly impacts treatment adherence, child health outcomes, and continuity of care. According to the AAP (2002), positive experiences in pediatric care benefit children's immediate

health and help establish lifelong trust and engagement with healthcare systems (COHC & IPFCC, 2012; Vaughn & Snively, 2024).

Patient experience has increasingly become a key measure of healthcare success, significantly influencing healthcare reimbursement. The shift from volume to value-based care has made patient experience even more important as an indicator of quality. For example, beginning in 2024, Medicaid programs must report standardized patient experience measures to keep receiving federal funding. This means patient experience is now directly tied to financial sustainability. Despite this critical role, few patient experience surveys have been validated specifically for pediatric outpatient care. Surveys currently used in pediatric settings are typically designed for adult populations or hospital settings and might not fully capture families' unique experiences in outpatient pediatric care.

Conceptual Framework and Measure Development

Despite the importance of pediatric primary care and the widespread use of patient experience surveys, a review of existing tools found that none currently used were developed with direct input from parents and caregivers in pediatric primary care. Closing this gap is crucial to improving relationship quality, promoting children's healthy development, and informing family-centered clinical practices.

Development of the CONNECT Survey

To address this need, a partnership of Reach Out and Read and ICS co-developed with parents a parent/caregiver-clinician relationship survey designed specifically for pediatric outpatient clinics. From the start, this process centered the voices and perspectives of families, with parent leaders included in the research team and engaging parents and caregivers directly through in-depth "empathy interviews" across five states (California, Florida, New Mexico, Pennsylvania, and New Jersey). Importantly, parent interviewers led these discussions, encouraging open, honest sharing of experience. Interviews involved families with children aged 0–5 years, representing a diverse demographic background (43% Black/African American, 20% White; 94% mothers, 6% fathers; 54% on Medicaid, 26% on private insurance).

During these interviews, parents shared what mattered most to them at pediatric visits and what encouraged continued visits to the same clinician. Key themes based on the frequency and importance parents assigned to each were identified by the research team. Prominent themes included:

- Being actively listened to and feeling genuinely attended to.

- Clear communication about their child's health.
- Being respected, valued, and treated without judgment.
- Being treated by clinicians with kindness, empathy, and genuine care.
- Being engaged in shared decision-making.
- Accessibility to clinicians and health information.
- Positive interactions with office staff and a welcoming clinic environment.

Over 50 potential survey items, developed directly from parents' own words, were refined through multiple iterative rounds of structured feedback using the Delphi method. Parent leaders were the experts whose feedback guided each revision, helping to ensure every question truly captured what matters most to parents and their real-life experiences. At the end of the Delphi process, the parent/caregiver group reached a consensus, resulting in a concise, parent-informed 13-item survey reflecting family priorities.

Description of the CONNECT Survey

The finalized CONNECT survey is a concise 13-item parent co-designed measure designed to capture the parent/caregiver-clinician relationship (Table 1). It is completed by parents/caregivers following a pediatric well-child visit. Each item is rated on a 4-point Likert-type scale, ranging from "Completely Disagree" to "Completely Agree."

- 1 – Completely Disagree
- 2 – Somewhat Disagree
- 3 – Somewhat Agree
- 4 – Completely Agree

Table 1: CONNECT Survey Items

	My child's doctor...
1.	... made me feel valued and respected.
2.	... made a connection with me and my child.
3.	... made me feel safe and comfortable.
4.	... did not judge me.
5.	... was kind and showed they cared.
6.	... gave me full attention.
7.	... did not seem rushed during the visit.
8.	... worked with me in making decisions about my child's health.
9.	... gave me the information needed to make the right decision for me and my child.
10.	... talked about my child's health in the way I understood.
11.	... addressed my questions and concerns.
12.	... gave me information on how to get in touch with them.
13.	The doctor's office staff were respectful and professional.

Methodology

Study Design and Data Collection

The validation of the CONNECT survey was conducted using two distinct datasets:

1. A primary dataset collected from pediatric clinics via the Mid-Atlantic Reach Out and Read network, used for psychometric validation, and
2. A secondary dataset collected from Florida Reach Out and Read clinics was used to assess test-retest reliability.

Primary Development Sample (Mid-Atlantic Reach Out and Read Network)

The main psychometric validation study was conducted across 131 pediatric clinics that deliver Reach Out and Read in North Carolina, South Carolina, Virginia, Washington, D.C., and Maryland. Data collection occurred over six weeks and was timed to coincide with well-child visits for children aged 0–5.

Caregivers were invited to complete the CONNECT survey immediately following their well-child visit. To ensure linguistic accessibility, the survey was offered in both English and Spanish. A total of 9,825 completed surveys were collected. These data were used to examine descriptive patterns, internal consistency reliability, inter-item correlations, construct validity, and measurement invariance.

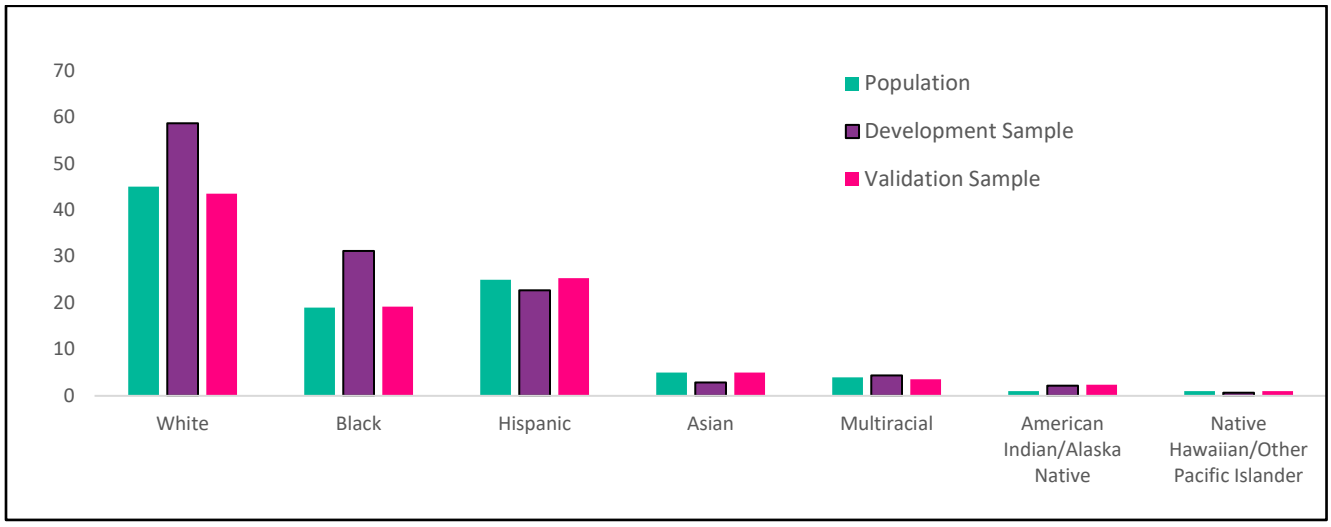
Characteristics of the Validation Sample

Although Reach Out and Read operates nationally, the development dataset was weighted toward sites in the Mid-Atlantic Reach Out and Read network. To enhance generalizability, we selected a validation subsample from the broader dataset that more closely aligns with the national demographics of Reach Out and Read clinics. The validation sample included urban, rural, and suburban clinics in North Carolina, South Carolina, Virginia, Maryland, and D.C. This subsample was used for all psychometric analyses. Figure 1 compares the population, development sample, and validation sample.

Demographic data for the validation sample showed a diverse group of respondents from across 127 Reach Out and Read sites. The majority were Medicaid-insured (58%), with others covered by private insurance (33%), Tricare (5%), or uninsured (2%). Regarding race and ethnicity, 44% identified as White, 19% as Black or African American, 2% as American Indian or Alaska Native, 5% as Asian, 4% as Multiracial, and less than 1% as Native Hawaiian or Pacific Islander. Twenty-five percent identified as Hispanic or Latino.

Most respondents indicated they were the primary caregiver bringing the child to medical visits and had previously seen the same provider. This continuity of care context provided an opportunity to evaluate how well the survey captured relational dynamics over time.

Figure 1: Comparison Between Population, Development Sample, and Validation Sample



Note. Population = ROR demographics; Development sample = Mid-Atlantic ROR demographics; Validation sample = Subset of development sample on which psychometric analyses are conducted.

Table 2: Descriptive Statistics of Mid-Atlantic ROR Validation Sample

During a typical week, how many days do you or another family member read to your child? (n = 2.953, mean = 3.84, SD = 1.08, Skewness = -0.59, Kurtosis = 2.43)		
	n	**%**
Never (1)	50	1.69
A few days (2)	353	11.95
Some days (3)	631	21.37
Most days (4)	902	30.55
Every day (5)	1015	34.37

Are you the one who typically takes the child to the doctor? (n = 2,963, mean = 3.63, SD = 0.62, Skewness = -1.56, Kurtosis = 2.43)		
	n	**%**
Never (1)	15	0.51
Sometimes (2)	180	6.07
Usually (3)	702	23.69
Always (4)	2066	69.73
How often have you seen this doctor before? (n = 2,949, mean = 3.63, SD = 0.62, Skewness = -1.56, Kurtosis = 4.46)		
	n	**%**
Never (1)	183	6.21
Not often (2)	184	6.24
A few times (3)	809	27.43
Many times (4)	1733	60.12
What type of health insurance does your child have?		
	n	**%**
None	59	1.99
Only Medicaid	1,723	58.09
Only Tricare	135	4.55
Only CHIP	7	0.24
Only Private	973	32.81

What is your race?		
	n	**%**
White, only	1,290	43.49
Black, only	570	19.22
Asian, only	149	5.02
Native Hawaiian or Other Pacific Islander, only	29	0.98
American Indian or Alaska Native, only	71	2.39
More than one race	107	3.61
Are you of Hispanic or Latino descent?		
	n	**%**
Hispanic or Latino	762	25.69
Non-Hispanic or Latino	2,204	74.31
How old is your child (months)? **(n = 2,966, mean = 22.24 months, SD = 19.65, Skewness = 0.90, Kurtosis = 2.60)**		
	n	**%**
Infants (0 - 12 months)	1,337	45.08
Toddlers (13 – 35 months)	851	28.69
Preschool (3 to 4 years)	482	16.25
Child (5 years)	296	9.98

Note. *n* = subgroup sample size.

Results

Survey Responses and Score Distribution

Descriptive statistics were calculated to examine patterns in caregiver responses across the 13 items of the CONNECT survey (Table 3). Each item was rated on a 4-point Likert scale, ranging from "Completely Disagree" (1) to "Completely Agree" (4).

The findings indicated uniformly high scores across all items. Mean item scores ranged from 3.92 to 3.95, with the overall average score across respondents being 3.94 (SD = 0.35). These consistently high values suggest strong agreement with statements reflecting positive parent/caregiver-clinician relationships.

In addition to mean scores, we assessed skewness and kurtosis to evaluate response distributions. Most items exhibited substantial negative skew and high kurtosis values, indicating that responses were concentrated at the upper end of the scale. For example, the item *"Talked about my child's health in the way I understood"* demonstrated a kurtosis value of 59.80, suggesting an overwhelming tendency for respondents to select the highest response category. The concentration of high scores across multiple items indicates a ceiling effect, which may limit the instrument's ability to distinguish between high-quality and exceptional relational experiences.

To assess convergent patterns, responses were also analyzed for two items drawn from the CAHPS Health Plan Survey and one from the Press Ganey Outpatient Medical Practice survey, which were included at the end of the CONNECT survey. These items, commonly used in health services research, also received high ratings:

- Likelihood to recommend the care provider (Press Ganey): Mean = 4.89 (out of 5)
- Doctor explained health in an understandable way (CAHPS): Mean = 3.89 (out of 4)
- Doctor showed respect for what the parent had to say (CAHPS): Mean = 3.95 (out of 4)

These results reinforce the broader trend of highly favorable caregiver-reported experiences within the participating clinics. However, they also highlight a potential limitation of the current scale in differentiating nuanced levels of relational quality, particularly in high-performing clinical environments.

Table 3: Descriptive Statistics of CONNECT Survey Items

My child's doctor...	n	Mean	Median	% Top Box	SD	Skewness	Kurtosis
...made me feel valued and respected.	2,963	3.93	4	96.05	0.37	6.69	50.15
...made a connection with me and my child.	2,965	3.92	4	94.54	0.39	5.90	40.94
...made me feel safe and comfortable.	2,965	3.94	4	96.59	0.36	7.08	55.30
...did not judge me.	2,966	3.94	4	96.49	0.37	6.95	53.22
...was kind and showed they cared.	2,965	3.94	4	96.76	0.37	7.11	55.20
...gave me full attention.	2,964	3.94	4	96.73	0.36	7.10	55.39
...did not seem rushed during the visit.	2,964	3.93	4	95.88	0.38	6.52	47.96
...worked with me in making decisions about my child's health.	2,964	3.94	4	96.52	0.36	7.08	55.34
...gave me information needed to make the right decision for me and my child.	2,963	3.94	4	96.49	0.36	7.10	55.87
...talked about my child's health in the way I understood.	2,964	3.95	4	97.03	0.35	7.41	59.80
...addressed my questions and concerns	2,964	3.94	4	96.79	0.36	7.26	57.76
...gave me information on how to get in touch with them.	2,965	3.93	4	95.82	0.38	6.48	47.18
Doctor's office staff were respectful and professional.	2,965	3.94	4	96.76	0.36	7.23	57.50
Average Score	2,962	3.94	4	92.07	0.35	7.35	59.45

My child's doctor...	n	Mean	Median	% Top Box	SD	Skewness	Kurtosis
Items from CAHPS Health Plan Survey							
How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?	2,964	3.89	4	91.30	0.42	4.58	27.05
How often did your child's personal doctor show respect for what you had to say?	2,961	3.95	4	95.54	0.23	5.31	35.42
Item from Press Ganey Outpatient Medical Practice Survey							
What is the likelihood of you recommending this care provider to others?	2,961	4.89	5	89.36	0.35	3.50	19.66

Note. Abbreviations: SD = standard deviation. % Top Box = percent of parents who selected the highest response category in the question.

Reliability of the CONNECT Survey

Internal Consistency Reliability

Internal consistency was assessed using Cronbach's alpha, McDonald's omega, and inter-item correlations. Cronbach's alpha estimates how consistently items measure the same concept, assuming equal item contributions. McDonald's omega accounts for differences in how strongly individual items relate to the underlying construct, and inter-item correlations examine how strongly individual items are related and detect any potential redundancy in item content.

Polychoric correlations are often preferred for ordinal Likert scale data analyses and were attempted for this analysis. However, due to severe skew and limited variability, internal consistency was calculated using Pearson correlations. To balance this limitation, we used multiple statistics to provide a more nuanced picture of the scale's performance including item-to-total (item-total) correlations to evaluate how well each item aligns with the overall scale score.

Cronbach's Alpha

The overall Cronbach's alpha was 0.993, indicating that the 13 survey items reliably measure a single coherent concept of relationship. The average inter-item covariance (0.123) suggests that survey items are closely related and reflect a shared construct.

At the item level:

- **Item-test correlations** ranged from 0.917 to 0.975, indicating strong alignment between each item and the overall construct.
- **Item-rest correlations** ranged from 0.901 to 0.971, showing no item detracted from the scale's reliability.

Item-test correlations reflect how strongly each item correlates with the scale's total score, indicating how well it represents the overall construct. Item-rest correlations show how well an item correlates with the total score, excluding that item, helping to identify whether any item reduces the scale's reliability.

While these findings support the scale's reliability, the exceptionally high alpha may also indicate item redundancy, meaning some items may be overly similar in what they measure.

Cronbach's Alpha by Racial/Ethnic Group

We assessed reliability within each subgroup to examine whether the scale performs consistently across racially and ethnically diverse populations (Table 4). Cronbach's alpha

remained high across all racial and ethnic subgroups (range: 0.981 to 0.994), confirming the scale's consistency across diverse populations. Average inter-item covariance was highest in the Black (0.184), Asian (0.166), and Multiracial (0.171) subgroups, and lowest in Native Hawaiian/Pacific Islander (0.035) and Native American/Alaska Native (0.042) subgroups. These differences may reflect how similarly participants responded to items or may be influenced by small subgroup sample sizes. Item-level reliability indicators were also strong across subgroups:

- **Black, only**: item-test mean = 0.970, item-rest mean = 0.964
- **Asian, only:** item-test mean = 0.966, item-rest mean = 0.957
- **Native Hawaiian / Pacific Islander, only**: item-test mean = 0.961, item-rest mean = 0.957
- **American Indian / Alaska Native, only**: item-test mean = 0.923, item-rest mean = 0.912
- **More than One Race**: item-test mean = 0.967, item-rest mean = 0.962
- **Hispanic participants**: item-test mean = 0.953, item-rest mean = 0.944

Table 4: Cronbach's Alpha by Racial and Ethnic Subgroup

Subgroup	Cronbach's Alpha (α)	Avg. Inter-item Covariance
Overall Sample	0.993	0.123
Black / African American	0.994	0.184
Hispanic / Latino	0.993	0.121
Asian	0.993	0.166
Native Hawaiian / Pacific Islander	0.989	0.035
Native American / Alaska Native	0.981	0.042
More than One Race	0.994	0.171

Cronbach's Alpha by Insurance Type

To explore whether internal consistency differs by insurance status, we analyzed reliability separately for participants with private and public insurance. As detailed in Table 5, internal consistency was similarly strong across insurance types.

Table 5: Cronbach's Alpha by Insurance Type

Subgroup	Cronbach's Alpha (α)	Avg. Inter-item Covariance
Private Insurance	0.9924	0.0997
Public Insurance	0.9931	0.1386

Item-level reliability indicators were also consistent:

- **Private insurance**: item-test mean = 0.961, item-rest mean = 0.953
- **Public insurance**: item-test mean = 0.965, item-rest mean = 0.957

Cronbach's Alpha by Age Group

We also examined whether the scale performs reliably across developmental stages by analyzing internal consistency by age group. Based on Cronbach's alpha, the parent–provider relationship scale demonstrated excellent internal consistency across all child age groups (Table 6).

Table 6: Cronbach's Alpha and Average Inter-item Covariance by Age Group

Age Group	Cronbach's Alpha (α)	Avg. Inter-item Covariance
Infants	0.993	0.115
Toddlers	0.985	0.074
Preschoolers	0.994	0.153
Children	0.998	0.253

These high values suggest reliable measurement across developmental stages. However, the high alphas in preschool and school-age children may again indicate item overlap.

To further evaluate item performance by age, we examined average item-test and item-rest correlations:

- **Infants**: item-test = 0.963, item-rest = 0.954
- **Toddlers**: item-test = 0.922, item-rest = 0.907

- **Preschoolers**: item-test = 0.968, item-rest = 0.962
- **Children**: item-test = 0.991, item-rest = 0.988

McDonald's omega

To complement Cronbach's alpha, we also calculated McDonald's omega, a reliability statistic that accounts for differences in how strongly each item contributes to the overall scale. Overall, the 13-item CONNECT survey scale yielded an omega coefficient of 0.993, indicating excellent internal consistency. This high value confirms that the items work together reliably to measure the intended construct. However, omega values this high may also suggest potential redundancy among items**,** meaning that some questions may be similar in what they measure.

McDonald's Omega Across Subgroups

To better understand how scale reliability varied across demographic groups, omega values were calculated where possible:

- **Black, only:** ω = 0.994
- **Hispanic participants:** ω = 0.992
- **Private insurance:** ω = 0.993
- **Public insurance:** ω = 0.993
- **Infants:** ω = 0.993
- **Toddlers:** ω = 0.986
- **Preschoolers:** ω = 0.994

Omega could not be calculated for several subgroups (Asian, Native Hawaiian/Pacific Islander, Native American/Alaska Native, Multiracial, and Children) due to model convergence issues. This is likely due to the limited variability in item responses in these subgroups, which makes it difficult for the model to generate stable estimates.

As with alpha, the high omega values observed across most subgroups support the scale's internal consistency reliability but suggest potential item redundancy.

Inter-Item and Item-Average Score Correlations

To further explore the internal structure of the instrument, inter-item correlations and item-to-total (item-average) correlations were examined. The average inter-item correlation was $r = .92$, and Spearman's correlation coefficients between item pairs ranged from $r = 0.63$ to $r = 0.89$. Polychoric correlations among items were consistently high ($r = .95$ to $.99$), (Tables 10 – 11, Appendix A), reflecting moderate to strong associations across all items, and the possibility of item redundancy. Correlations can be interpreted as follows: 0.6 or greater = strong, 0.4 – 0.59 = moderate, 0.2 – 0.39 = acceptable, less than 0.2 = weak (ICS, 2002)

These values support the interpretation that the survey items are conceptually related, functioning together to assess a common latent construct. However, when combined with the high alpha and omega values, the upper range of the correlations suggests that some items may capture overlapping content domains.

Summary

The CONNECT survey demonstrates exceptionally high internal consistency across the sample and subgroups: Cronbach's alpha and McDonald's omega exceeding 0.98. Item-level correlations provide evidence of strong alignment with the overall construct, confirming that the items are a reliable measure of relationship quality between parents and their child's provider.

While strong, these findings are limited by the variability in item-level responses, with most caregivers selecting the highest possible ratings (top-box scores). While common among measures of healthcare quality, this pattern may indicate item redundancy, sampling patterns, and/or formatting-related limitations. These are discussed more fully in the Discussion and Implications section of this document.

Test-Retest Reliability

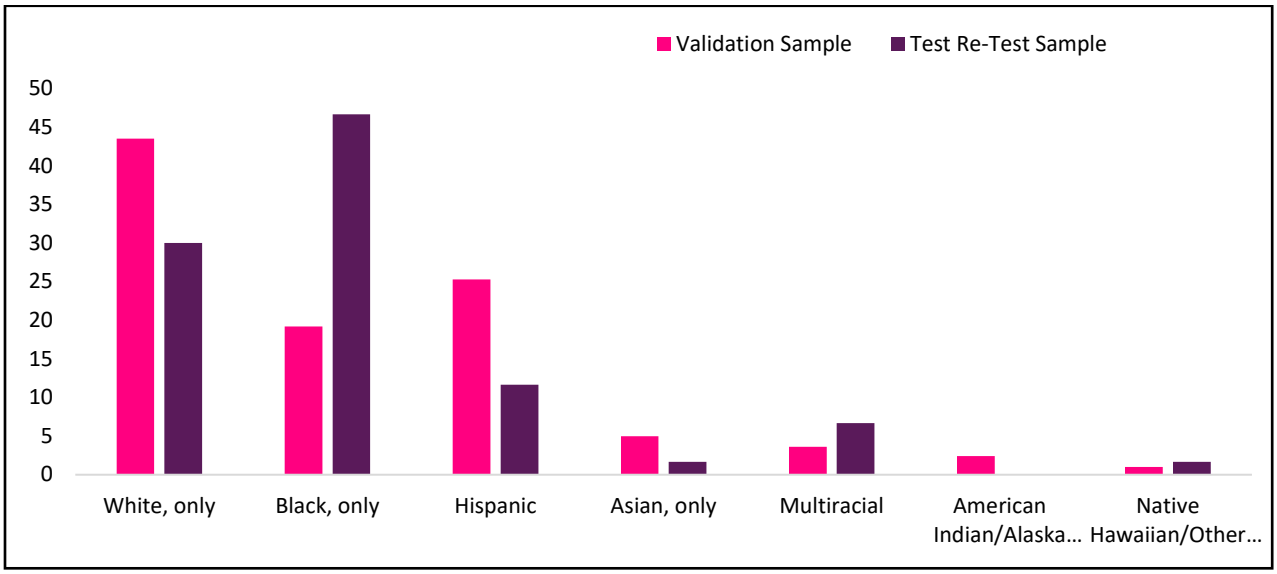
Test-Retest Reliability Subsample (Florida)

To assess the temporal stability of the CONNECT survey, a separate test-retest study was conducted in Florida using a different set of clinics that delivery the Reach Out and Read model. This component focused solely on test-retest reliability and was not included in the primary psychometric analyses.

Caregivers were asked to complete the CONNECT survey twice—immediately following their child's well-child visit and again within 5 to 7 days of the initial administration. Responses were matched across time points using unique identifiers, resulting in a final analytic sample of 60 matched parent/caregiver responses.

Figure 2 shows the demographic characteristics of the test-retest sample compared to the validation sample. This dataset was used exclusively to assess response consistency over time and is presented separately from the primary validation findings to ensure analytic clarity.

Figure 2: Comparison Between Validation Sample and Test-Retest Sample



As with the validation sample, the CONNECT survey test-retest sample presented substantial ceiling effects. Across test and retest items, between 80% and 93.3% of participants selected the highest response item (top-box score), resulting in limited variability across test times.[1]

[1] As a result, conventional test-retest analyses, such as polychoric correlations and confirmatory factor analysis (CFA), could not be computed or yielded unstable results. The limited response variability also undermines the interpretability of traditional reliability metrics, such as Cronbach's alpha and McDonald's omega, as high estimates in this context may reflect item redundancy rather than true internal consistency.

Test–retest reliability was evaluated using Gwet's AC1 for item-level analysis to accommodate the ordinal nature of the items, skewed distributions, and high inter-item redundancy. Unlike Cohen's kappa, which can underestimate agreement when there is limited response variability, AC1 is more stable and reliable in contexts where ratings are highly concordant or clustered within a particular response category. This makes it particularly appropriate for evaluating test-retest reliability when the data exhibit ceiling effects or limited dispersion across categories. Gwet's AC1 statistics are interpreted as follows: poor (< 0), slight (0 – 0.2), fair (0.21 – 0.40), moderate (0.41 – 0.60), substantial (0.61 – 0.80), and almost perfect (0.81-1.00) (Landis & Koch, 1977).

Percent agreement was also calculated across items. Given the test-to-retest time frame (5-7 days), a percent agreement of at least 80% is considered acceptable, and agreement over 90% is excellent (ICS, 2022). Due to limited variability, percent agreement should be interpreted with caution.

For scale-level reliability, in addition to percent agreement and Gwet's AC1, Spearman's rank correlation coefficient and polychoric correlations were calculated between timepoints to assess rank-order consistency of summed scores. However, extreme ceiling effects and limited item variability may distort the estimates.

Test-Retest Results

Results demonstrated high levels of agreement across individual items and the overall scale (see Table 7).

Table 7: Test-Retest Reliability

Item	% agreement	Gwet's AC1
Average Score	95.15	0.92
My child's doctor…		
… made me feel valued and respected.	94.05	0.93
… made a connection with me and my child.	90.00	0.87
… made me feel safe and comfortable.	94.05	0.93

Item	% agreement	Gwet's AC1
… did not judge me.	95.54	0.95
… was kind and showed they cared.	94.05	0.93
… gave me full attention.	92.26	0.91
… did not seem rushed during the visit.	94.05	0.93
… worked with me in making decisions about my child's health.	94.05	0.93
… gave me the information needed to make the right decision for me and my child.	94.05	0.93
… talked about my child's health in the way I understood.	95.24	0.94
… addressed my questions and concerns.	94.94	0.94
… gave me information on how to get in touch with them.	94.05	0.92
…The doctor's office staff were respectful and professional.	94.05	0.93

Note. n = 56.

Across items, agreement ranged from 90% to 95.54%, with corresponding Gwet's AC1 coefficients from 0.87 to 0.95. The survey average score, calculated as the mean of all items, showed the highest consistency over time:

- **Percent agreement =** 95.15%
- **Gwet's AC1 =** 0.92
- **Spearman's rho =** 0.72, $p < .001$
- **Polychoric** r = 0.45, SE = 0.11

The difference in agreement between the average score and individual items is likely due to increased response variability at the scale level, where small item-level differences are smoothed out through averaging. These findings suggest strong overall test–retest stability of the CONNECT survey. However, due to the low variability and high redundancy among items, this data should be interpreted with caution.

Summary

A test-retest study was conducted with 56 caregivers who completed the CONNECT survey twice within 5 to 7 days. Results showed high agreement across items (90–95.5%) and strong reliability, with Gwet's AC1 values ranging from 0.87 to 0.95, and a Spearman's rho of 0.72 and polychoric *r* of 0.45 for the total score. While these findings support the measure's consistency over time, ceiling effects and limited response variability may inflate agreement estimates and mask meaningful differences. The high redundancy among items also suggests potential overlap in content.

Validity of the CONNECT Survey

Construct Validity (Factor Analysis)

Construct validity refers to the extent to which the instrument accurately measures the theoretical concept it intends to assess—in this case, the quality of the parent/caregiver-clinician relationship. Although polychoric correlation is typically recommended for factor analysis of ordinal data, attempts to generate a polychoric matrix resulted in a non–positive definite matrix due to extremely high inter-item correlations and limited response variability. As a result, to evaluate the internal structure of the CONNECT survey, we conducted an exploratory factor analysis (EFA) using Pearson correlations with principal factor extraction and Promax rotation.

The initial unrotated solution yielded six factors with eigenvalues above 1. However, the first factor had a substantially higher eigenvalue (11.93), accounting for 98.93% of the total variance. All subsequent factors had eigenvalues below 0.20, contributing minimal additional explanatory value. To determine the appropriate number of factors to retain, we performed a parallel analysis, which compares the observed eigenvalues to those derived from randomly generated data. Both scree plots exhibited a steep decline after the first component, reinforcing the presence of a dominant underlying factor and supporting the unidimensional structure of the scale (Figures 3 and 4).

Figure 3: EFA Scree Plot Showing One Factor Model

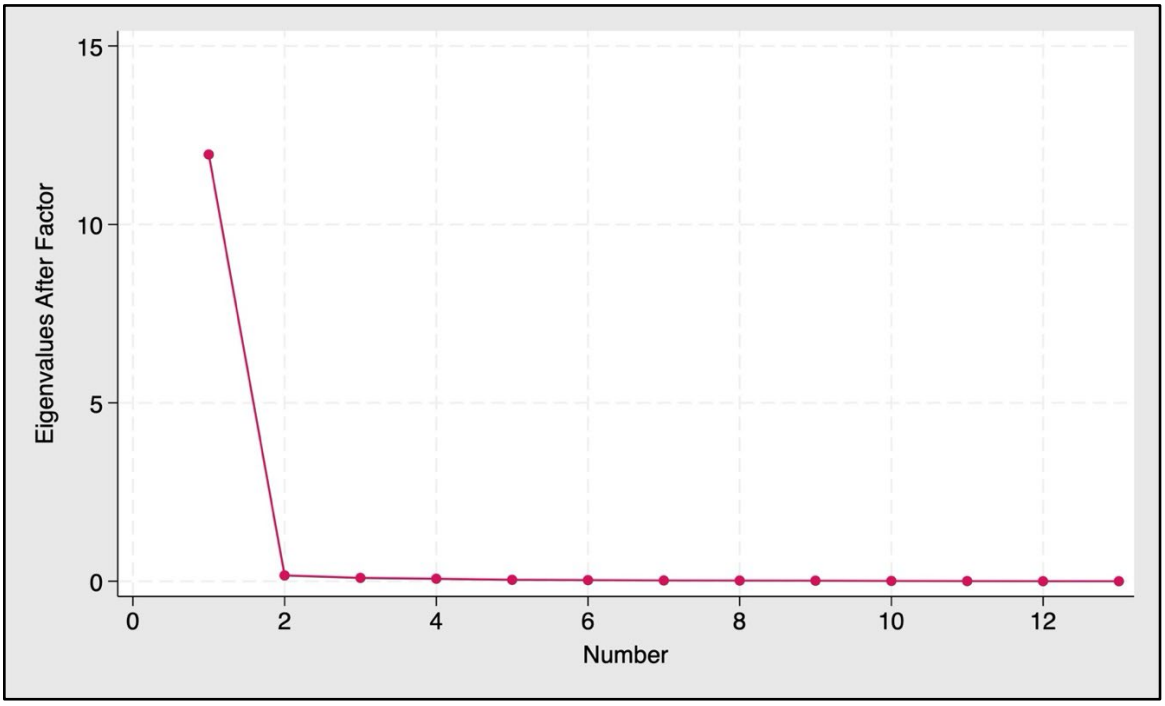
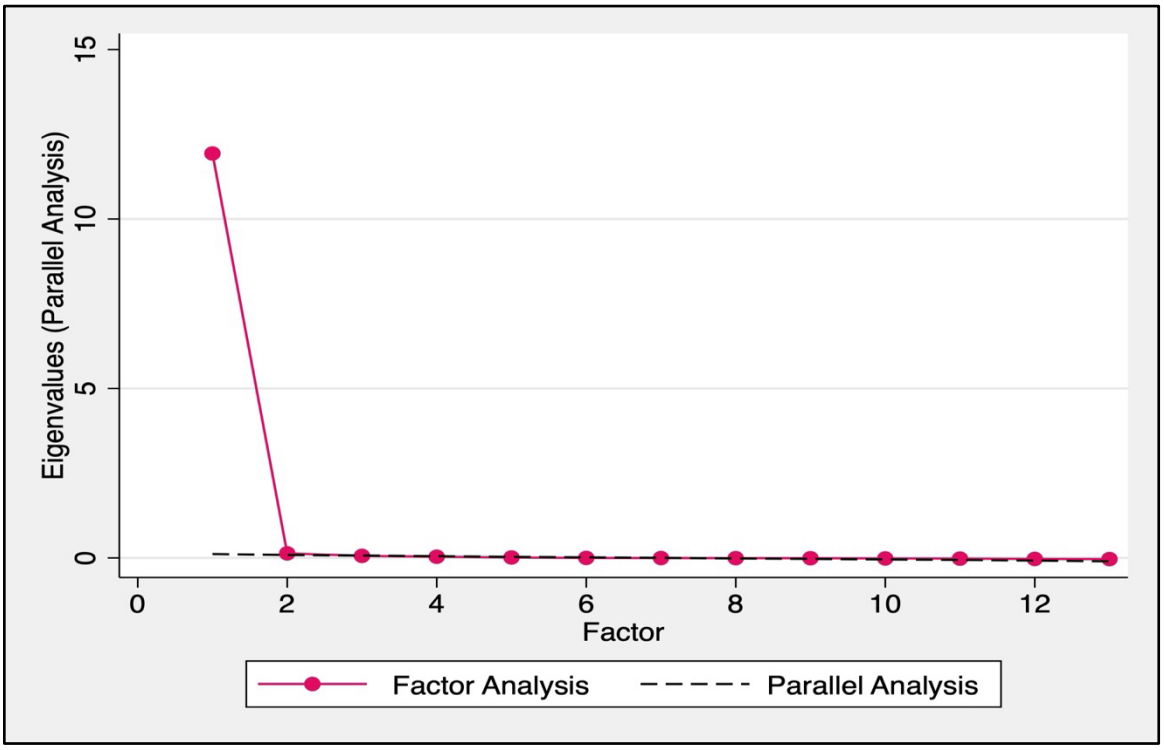


Figure 4: Parallel Analysis Scree Plot Showing One Factor Model



We applied an oblique Promax rotation to explore potential secondary dimensions and conducted a forced two-factor solution. Rotation revealed that all items loaded strongly on the first factor with limited and inconsistent loadings on the second (Table 8). Furthermore, the two rotated factors were highly correlated (r = 0.94), reinforcing the presence of a single dominant latent construct.

Factor loadings for the primary factor ranged from 0.66 to 0.73 for most items, and uniqueness values (i.e., the proportion of variance not explained by the factor solution) were low for all but one item. This pattern further supports a unidimensional structure while suggesting that minor secondary dimensions may exist but are not distinct enough to warrant separate factors. Attempts to conduct confirmatory factor analysis (CFA) using both structural equation modeling and generalized structural equation modeling encountered convergence and identification issues, likely due to extreme ceiling effects and multicollinearity among items. This pattern is consistent with the results of the exploratory factor analysis and high internal consistency (α = .99, ω = .99), supporting a unidimensional structure despite statistical challenges to model estimation. These results support the conclusion that the CONNECT survey captures a single, unified concept of relationship quality.

Table 8: Factor Loadings > 0.30 of the CONNECT Survey Items in a 2-Factor Forced Rotation

Item	Factor 1 Loading	Factor 2 Loading	Uniqueness
My child's doctor...			
...made me feel valued and respected.	0.325	0.693	0.060
...made a connection with me and my child.		0.693	0.159
...made me feel safe and comfortable.	0.439	0.591	0.048
...did not judge me.	0.330	0.694	0.049
...was kind and showed they cared.	0.396	0.627	0.060
...gave me full attention.	0.537	0.489	0.061
...did not seem rushed during the visit.	0.415	0.581	0.112

Item	Factor 1 Loading	Factor 2 Loading	Uniqueness
My child's doctor...			
...worked with me in making decisions about my child's health.	0.549	0.473	0.067
...gave me the information needed to make the right decision for me and my child.	0.675	0.356	0.040
...talked about my child's health in the way I understood.	0.712	0.320	0.033
...addressed my questions and concerns.	0.703	0.323	0.044
...gave me information on how to get in touch with them.	0.658	0.313	0.145
The doctor's office staff were respectful and professional.	0.725		0.055

Note. Uniqueness = Proportion of that item's variance that is not explained by the underlying factor. A high uniqueness value suggests that most of a variable's variance is unique to that variable. In contrast, a low uniqueness value indicates the variable is primarily explained by the factors extracted in the analysis.

Convergent Validity

Convergent validity describes how well a measure compares to established measures of the same construct. It provides evidence that a measure accurately reflects the intended concept(s) and contributes to overall measure validity (ICS, 2022).

Press Ganey and CAHPS

To assess convergent validity, we examined associations between the CONNECT survey and caregiver-reported items collected at the same time from two widely used measures of healthcare quality: the Consumer Assessment of Healthcare Providers and Systems (CAHPS) (Agency for Healthcare Research and Quality, 2023) survey and the Press Ganey® Outpatient Medical Practice Survey (n. d). Two items from the CONNECT survey were compared to corresponding items from the CAHPS survey, a nationally standardized tool used to measure

patients' experiences with healthcare providers and services. CAHPS items were selected because they reflected communication and relational behaviors that we predicted would correlate with specific CONNECT items. In addition, we compared the CONNECT average score to the "likelihood to recommend" item from the Press Ganey survey, which is used as a proxy for overall patient experience. The Press Ganey survey is widely used to measure patient and caregiver experiences in healthcare settings (North & Tulledge-Scheitel, 2019). Specifically, we compared the following:

Measure	Comparison Item	CONNECT Item
CAHPS	How often did your child's personal doctor explain things about your child's health in a way that was easy to understand? *	My child's doctor talked about my child's health in the way I understood. (Item 10) *
CAHPS	How often did your child's personal doctor show respect for what you had to say? *	My child's doctor made me feel valued and respected. (Item 1) *
Press Ganey	What is the likelihood of you recommending this care provider to others? **	Average CONNECT score

Note. *Item measured on a 4-point Likert scale. **Item measured on a 5-point Likert scale

Findings

Correlations

Spearman correlations were computed between these external measures, corresponding CONNECT survey items, and overall scores. This approach is appropriate for ordinal data, such as Likert scales, data with high ceiling effects, and comparing items across scales with different responses (i.e., 4-point vs. 5-point Likert scales). By ranking the data, Spearman's rho minimizes the influence of differing scale ranges and focuses on the strength and direction of the relationship between items and/or measures. As is preferable for ordinal Likert items, Polychoric correlations were also calculated between the CAHPS and CONNECT items but were not appropriate for comparisons to the CONNECT average score. Correlations can be interpreted as follows: 0.6 or greater = strong, 0.4 – 0.59 = moderate, 0.2 – 0.39 = acceptable, less than 0.2 = weak (ICS, 2002). Results indicated acceptable but statistically significant associations:

1. **Overall scores:** CONNECT survey average score and Press Ganey *likelihood to recommend* (*rho* = 0.29, $p < .001$).
2. **Communication items:** CONNECT *"My child's doctor talked about my child's health in the way I understood"* and CAHPS *"How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?"* (*rho* = 0.23, $p < .001$) ($r_p = 0.60$, SE = 0.05)
3. **Respect items:** CONNECT *"My child's doctor made me feel valued and respected,"* and CAHPS *"How often did your child's personal doctor show respect for what you had to say?"* (*rho* = 0.30, $p < .001$) ($r_p = 0.50$, SE = 0.05)

Regression Models

Polychoric correlations between CONNECT and CAHPS items suggested that the underlying assumption of bivariate normality may not have been met (respect items: Pearson chi-square (G^2) = 46.36, *df* = 8, $p < .001$, and a likelihood-ratio chi-square (LR χ^2) = 50.83, *df* = 8, $p < .001$) (communication items: Pearson chi-square (G^2) = 48.76, *df* = 8, $p < .001$, and a likelihood-ratio chi-square (LR χ^2) = 67.96, *df* = 8, $p < .001$). In addition, as a quasi-ordinal variable, the CONNECT average score does not meet the requirements for polychoric correlation. As a result, we employed a two-part model using binary logistic regression to examine the odds of a "top-box" CONNECT and CAHPS/Press Ganey score. Next, we used ordered logistic regression to analyze the relationship between less-than-perfect CONNECT scores and their corresponding CAHPS and Press Ganey items. These approaches addressed issues with limited variability and allowed for more accurate estimation of convergent validity.

Results showed a statistically significant association between the CONNECT top-box average score and the top-box response to the Press Ganey "likelihood to recommend" item. (OR = 8.02, SE = 1.19, $p < .001$, 95% CI [5.99 - 10.74] pseudo $R^2 = 0.08$). Caregivers who selected the highest possible rating on the CONNECT scale had over 8 times the odds of also selecting the highest rating on the Press Ganey "likelihood to recommend question." The relationship between less-than-perfect CONNECT and Press Ganey scores was also statistically significant (OR = 2.11, SE = 0.74, $p = 0.03$, 95% CI [1.06 - 4.18], pseudo $R^2 = 0.06$). Even among less satisfied respondents, caregivers who selected a higher CONNECT rating had twice the odds of selecting a higher rating

on the Press Ganey "likelihood to recommend" item, indicating a strong relationship between the two measures.

The relationship between the scores on the CONNECT and CAHPS respect items was also statistically significant. Caregivers who selected the "top-box" score on the CONNECT respect item had 16.35 times the odds of also selecting the "top-box" score on the CAHPS respect item (OR = 16.35, SE = 3.55, $p < .001$, 95% CI [10.68 - 25.04, pseudo $R^2 = 0.11$). Among less satisfied caregivers, each one-point increase on the CONNECT respect item was associated with 8.77 times greater odds of selecting a higher rating on the CAHPS respect item (OR = 8.77, SE = 7.84, $p = 0.02$ 95% CI [1.52 - 50.54], pseudo $R^2 = 0.41$).

On the communication items, caregivers that selected the highest rating on the CONNECT communication item had 11.21 times the odds of also selecting the highest rating on the CAHPS communication item (OR = 11.21, SE = 2.49, $p < .001$, 95% CI [7.25 - 17.34, pseudo $R^2 = 0.06$). Among participants who did not select "top-box" responses, there was no statistically significant association between CONNECT and CAHPS communication scores (OR = 2.18, SE = 1.06, $p = 0.11$, 95% CI [0.83 - 5.68, pseudo $R^2 = 0.06$).

While results of correlation analysis demonstrated acceptable to strong relationships between CAHPS and CONNECT items, the two-part regression model provided stronger evidence of association and better model specification, suggesting a much stronger relationship than correlation coefficients alone. These offer stronger evidence of the CONNECT's convergent validity while also accounting for the ordinal nature of the data and avoiding distributional violations that impact the polychoric correlations.

Figures 4, 5, and 6 provide both parametric and non-parametric visualizations of the relationship between the CONNECT, CAHPS, and Press Ganey scores and items. Points are jittered to reduce overlap due to high clustering at top response values. LOWESS smoothing curves and linear prediction lines were used to illustrate the observed relationship and model-based trends between items.

Figure 5: Relationship Between CONNECT Average Score and Press Ganey "Likelihood to Recommend"

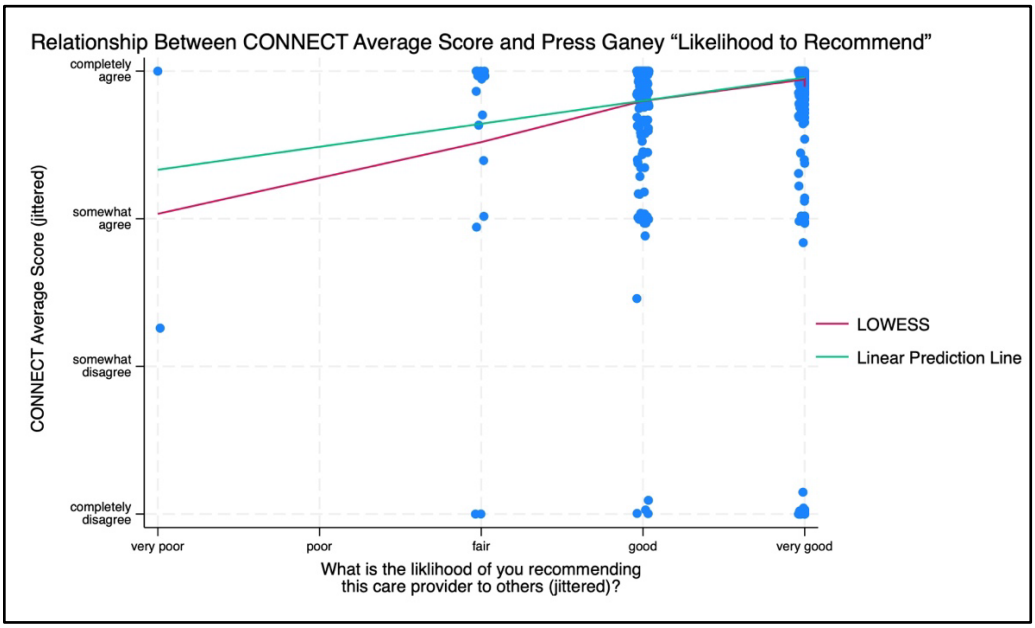


Figure 6: Relationship Between CONNECT & CAHPS Communication Items

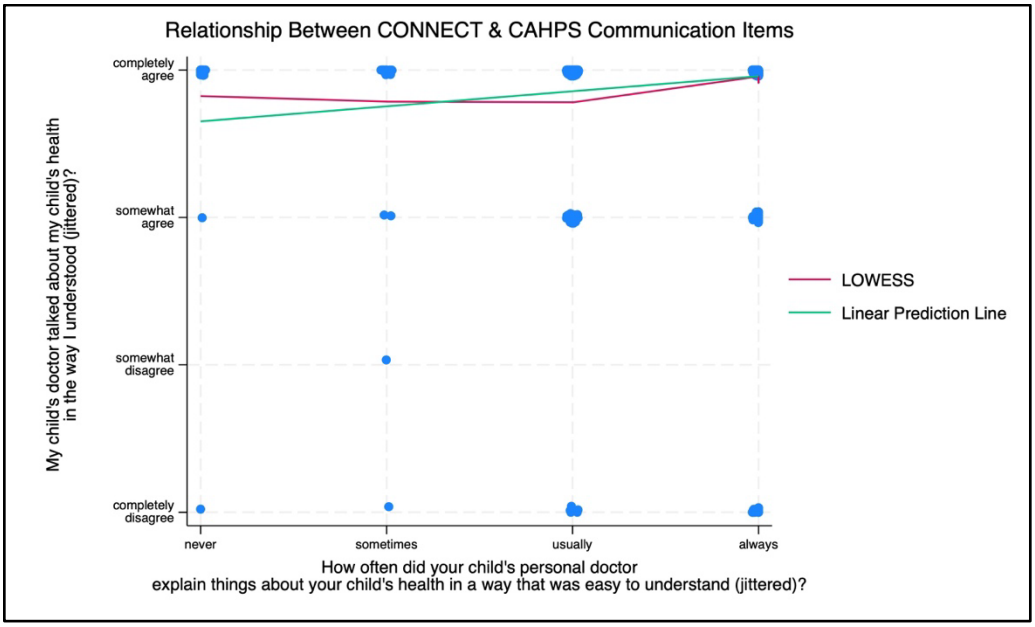
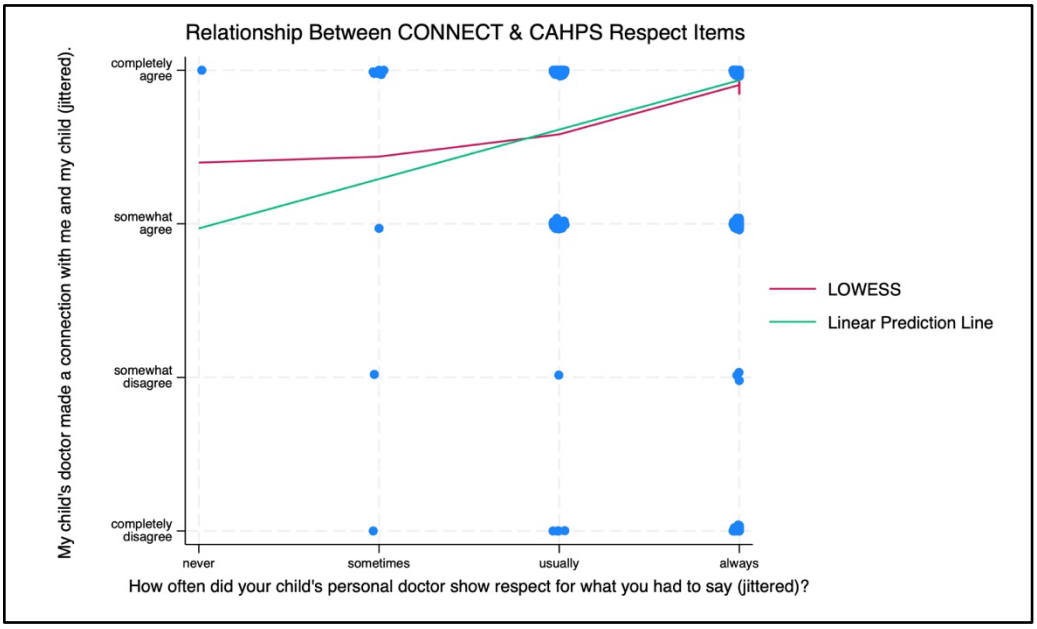


Figure 7: Relationship Between CONNECT & CAHPS Respect Items



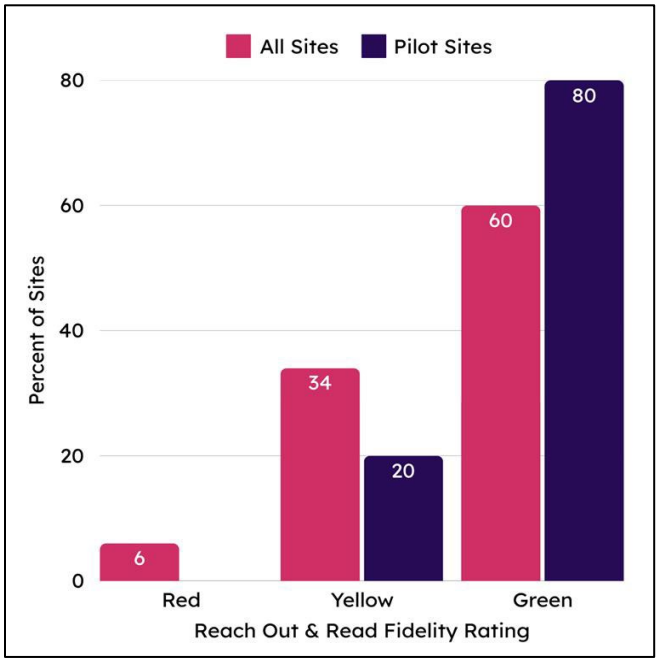
Reach Out and Read Site Quality Classifications

Reach Out and Read (2024) uses Site Quality Classifications (SQC) to enhance the quality of each site's implementation and identify target areas for improvement. The classification score measures 14 core components of high-quality Reach Out and Read sites in the following domains:

1. **Medical Providers:** Four items, including the percentage of trained providers, provider advocacy, and whether providers provide a book and anticipatory guidance at the beginning of well-child visits.
2. **Books:** Four indicators, including a system for tracking book distribution to providers and children, and book quality.
3. **Literacy Rich Environment:** Includes two items assessing whether the waiting and/or exam rooms promote literacy and whether the site provides information about community-based literacy resources such as libraries.
4. **Program Management:** Four items that measure book funding, tracking systems (demographic, well-child visits) and progress reports, supportive leadership, and the presence and accessibility of a site coordinator.

One point is awarded per item, with site classifications awarded as follows: 14 points = green classification, 11 – 13 points = yellow classification, and 10 points or less = red classification. Figure 7 compares the site classification ratings between the pilot and all Reach Out and Read sites.

Figure 8: Reach Out and Read SQC Score Comparison (Pilot vs. All ROR sites)



The Reach Out and Read SQC was used as an additional measure of convergent validity. We theorized that sites with higher SQC scores would also demonstrate slightly higher average scores on the CONNECT survey, despite the difference in each measure's intended construct. Given the difference in constructs between the CONNECT and SQC, we recognized that the relationship would likely be moderated by demographic characteristics of Reach Out and Read participants at each site (Gotschall, et al., 2024; Haviland, et al., 2005; Johnson et al., 2024; Murray-Garcia, et al., 2000; Turpin et al., 2021).

Findings

Correlations

The correlation between CONNECT average scores and SQC classifications were weak and insignificant (Spearman's $r = 0.03$, polychoric $r = 0.04$). These findings were unsurprising given the difference in constructs and were used to inform subsequent analysis between the two measures.

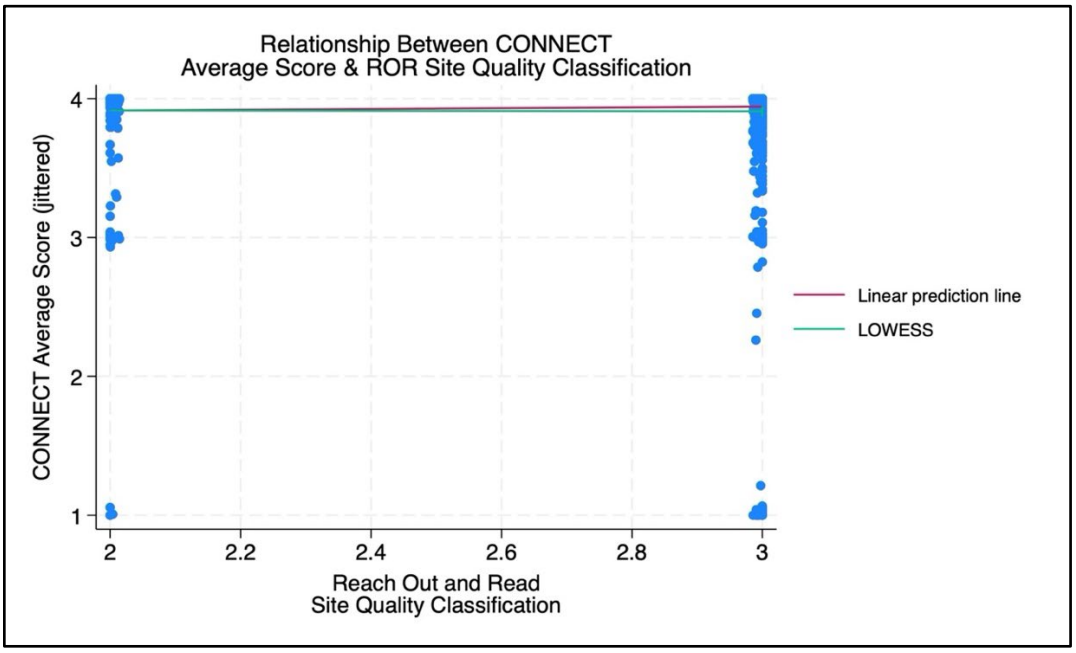
Regression Models

Ordinal logistic regression was used to assess convergent validity by examining whether CONNECT survey scores varied in expected ways across site classification levels (Figure 9). Due to the limited variability in site classification scores, "top-box" models without controls could not be used in this analysis. Total SQC points and CONNECT average scores were utilized to complete "less-than-perfect" models without controls. Results indicate that, each one-point increase in SQC points was associated with 25% higher odds of reporting a higher CONNECT average score (OR = 1.25, SE = 0.14, LR $\chi^2(1) = 3.62$, $p = 0.05$, 95% CI [1.00 - 1.55], pseudo $R^2 = 0.001$).

In addition, with control variables (Wald $\chi^2(13) = 45.54$, $p < 0.001$, pseudo $R^2 = 0.017$), patients at green-classified sites had 47% higher odds of reporting higher scores on the CONNECT survey (OR = 1.47, $p = 0.039$), consistent with the known quality distinctions between sites. A parallel pattern was observed for Press Ganey scores (Wals $\chi^2(13) = 68.78$, $p < 0.001$, Pseudo $R^2 = 0.029$). Patients at green-classified sites had 45% higher odds of reporting a higher rating on the "likelihood to recommend" score than those at yellow-classified sites (OR = 1.45, $p = 0.028$). These findings further support the convergent validity of the CONNECT survey.

Although the odds ratios indicate statistically significant relationships, the effect sizes were modest. These findings suggest that CONNECT scores may reflect factors beyond what is captured in the SQC rubric. We anticipated this result as the CONNECT and Reach Out and Read SQC measure unique, but overlapping characteristics of Reach Out and Read sites.

Figure 9: Relationship Between CONNECT Average Score & Reach Out and Read Site Quality Classification



Concurrent Validity

Concurrent validity evaluates how well a measure correlates with characteristics of the measured population. To establish concurrent validity, we examined the relationship between the CONNECT survey average score the frequency of visits with the same provider and whether the caregiver is the primary adult responsible for bringing the child to the doctor.

Findings

Correlations

As anticipated, both Spearman and polychoric models demonstrated weak, but statistically significant correlations between frequency of visits with the same provider and whether the caregiver that participated in the study is typically the adult that brings the child to visits. At the item-level, the relationship between variables and provider frequency were weak (Spearman's r = 0.06 - r = .09; polychoric r = 0.13 - r = 0.18). Item-level findings for caregiver role were similar (Spearman's r = 0.06 - r = .11; polychoric r = 0.17 - r = 0.22). The relationship between

the CONNECT average score and frequency of visits with the same provider (Spearman's r = .09; polychoric r = 0.06) and caregiver role (Spearman's r = .07; polychoric r = 0.05).

Regression Models

Once again, we used a two-part model. Binary logistic regression was employed to examine the odds of a "top-box" CONNECT average score and the highest frequency response for caregiver role (always) and provider frequency (many times). Then, we used ordered logistic regression to analyze the relationship between less-than-perfect CONNECT scores and lower response categories across caregiver/provider items. Due to the weak, but statistically significant correlation between caregiver role and frequency of visits with the same provider (Spearman's r = 0.11, tetrachoric r = 0.18) regression models that included all three variables were also completed.

Caregivers that selected that they were "always" the individual that brings the child to the doctor were 58% more likely to report a "top-box" CONNECT average score (OR = 1.58, SE = 0.22, p = 0.001, 95% CI [1.29 - 2.07], LR $\chi^2(1)$ = 10.40, pseudo R^2 = 0.006). In addition, caregivers that selected that they have seen this provider "many times" had 90% higher odds of reporting a "top-box" CONNECT average score (OR = 1.90, SE = 0.26, LR $\chi^2(1)$ = 22.51, p < 0.001, 95% CI [1.45 - 2.48], pseudo R^2 = 0.01). When all three variables were included in the model (pseudo R^2 = 0.02), caregivers that were "always" the individual that brings the child to the doctor had 47% higher odds of reporting a "top-box" CONNECT average score (OR = 1.47, SE = 0.21, LR $\chi^2(1)$ = 29.95, p = 0.006, 95% CI [1.12 - 1.94]). Those that had seen the provider "many times" had 83% higher odds of reporting a "top-box" CONNECT average score (OR = 1.83, SE = 0.25, LR $\chi^2(1)$ = 29.95, p < 0.001, 95% CI [1.39 - 2.39]).

There was also a statistically significant relationship between "less-than-perfect" CONNECT average scores and if the caregiver is typically responsible for bringing the child to the doctor (OR = 2.63, SE = 1.09, p = 0.02 1, 95% CI [1.17 - 5.94], LR $\chi^2(1)$ = 5.51, pseudo R^2 = 0.01). Caregivers that "usually" brought the child to the doctor had 163% higher odds of reporting higher CONNECT scores. There was no statistically significant relationship between CONNECT average score and the frequency of visits with the same provider (OR = 0.77, SE = 0.18, p = 0.26, 95% CI [0.47 - 1.21], LR $\chi^2(1)$ = 1.30, pseudo R^2 = 0.01). When all three variables were included in the model (LR $\chi^2(2)$ = 8.89, pseudo R^2 = 0.03), caregivers that brought the child to the doctor had 171% higher

odds of reporting a higher CONNECT average score (OR = 2.72, SE = 1.39, *p* = 0.05, 95% CI [1.00 - 7.38]) and caregivers that see the same doctor less frequently had 62% lower odds of reporting a higher CONNECT average score (OR = 0.38, SE = 0.15, *p* = 0.02, 95% CI [0.17- 0.84].

These findings provide evidence of concurrent validity for the CONNECT survey. While correlations were modest, the consistent and statistically significant associations between frequency of visit with the same provider, caregiver role, and CONNECT scores provide additional evidence that the CONNECT captures meaningful differences in the caregiver-provider relationship. These results support the utility of the CONNECT survey as a valid tool for capturing caregiver perspectives, particularly when used in conjunction with contextual information about care delivery. Figures 10 and 11 depict the parametric and non-parametric relationship between these variables.

Figure 10. Relationship Between CONNECT Average Score & How Often Caregiver Brings Child to Doctor

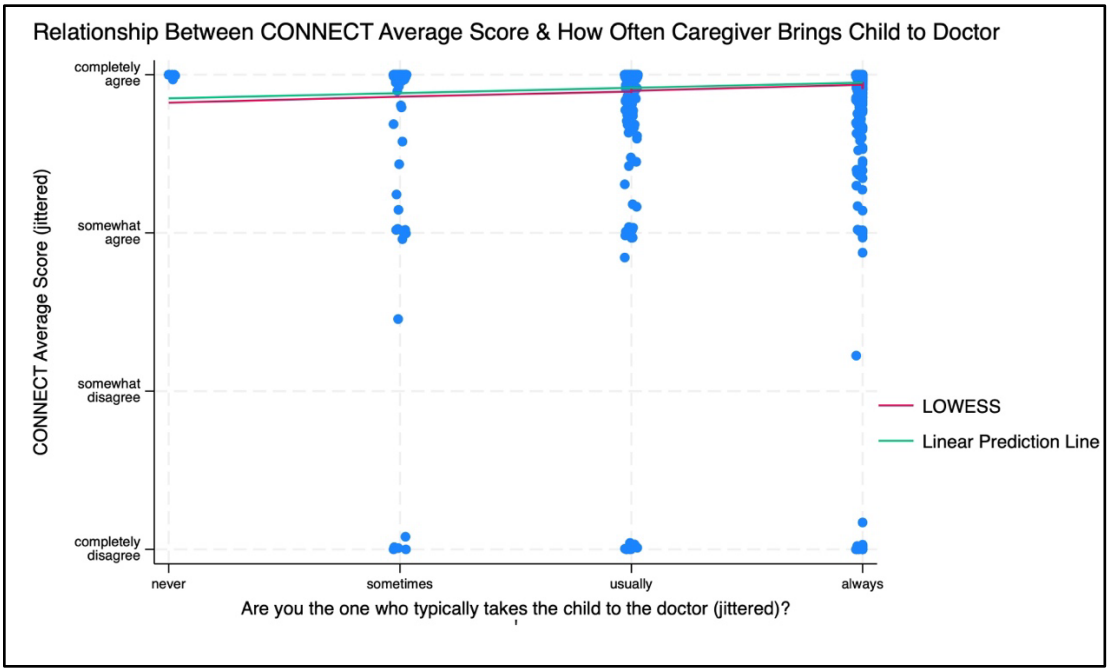
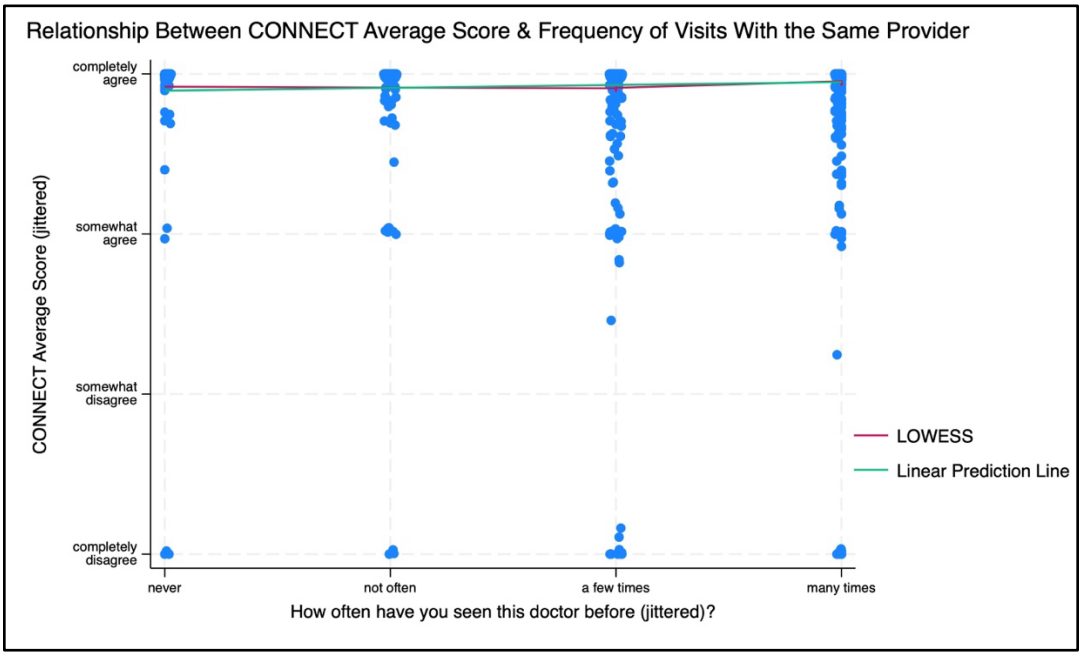


Figure 11. Relationship Between CONNECT Average Score & Frequency of Visits with the Same Provider



Summary

Evidence of construct validity was established through EFA, confirming that the CONNECT survey measures a single underlying construct: the caregiver–provider relationship quality. A dominant first factor explained over 98% of the variance, and item loadings were consistently strong, supporting the scale's unidimensional structure. Convergent validity was demonstrated through statistically significant correlations between the CONNECT survey and items from the CAHPS and Press Ganey surveys, including caregiver-reported respect, communication, and likelihood to recommend. While correlations were acceptable (r = 0.23–0.30), they were directionally aligned and consistent with theoretical expectations. Convergent validity was supported by findings from ordinal logistic regression models. Caregivers at green-classified Reach Out and Read sites were significantly more likely to report higher CONNECT survey scores (OR = 1.47, p = 0.039) and higher "likelihood to recommend" ratings (OR = 1.45, p = 0.028). These findings indicate good overall validity of the CONNECT survey. Concurrent validity analyses showed that caregivers who consistently see the same provider and are primarily responsible for the child's visits were more likely to report top CONNECT scores. Though correlations were modest,

regression models confirmed statistically significant relationships, supporting the CONNECT survey's ability to capture meaningful relational dynamics in pediatric care.

Measurement Invariance

Analytic Approach

To check whether the CONNECT survey works equally well across different demographic groups, we performed statistical tests to evaluate systematic differences in responses. Our goal was to determine if parents/caregivers from different backgrounds answered the questions differently, which might suggest bias in the survey or different experiences.

Wilcoxon Rank-Sum Tests

Since the survey responses are ordinal (ranked) and were not normally distributed, we used a three-step approach to analyze this data. First, we used Wilcoxon rank-sum tests to see if responses differed significantly between demographic groups when categorized into two groups. This non-parametric test is well-suited for non-normally distributed data and provides an initial assessment of group-level differences in overall CONNECT survey scores and item-level responses. Binary comparisons were created for the following variables:

1. *Insurance type* (private vs. all others, public vs. all others, self-pay/no insurance vs. all others)
2. *Race/Ethnicity (*White vs. all others, Black vs. all others, Asian vs. all others, American Indian/Alaska Native vs. all others, Multiracial vs. all others, Hispanic/Latino vs. all others)
3. *Child age group* (infants/<=12 months vs. all others, toddlers/13-35 months vs. all others, preschool/36-59 months vs. all others, children/60-71 months vs. all others)

Table 9 presents the number of participants within each binary-coded group used in these analyses.

Table 9: Number of Participants per Binary Subgroup Variable

Variable	*n*
Binary Race/Ethnicity Variables	
Black or African American	570
White	1,290
Native American or Alaska Native	71
Asian	149
Native Hawaiian or Other Pacific Islander	29
More than one race	107
Hispanic or Latino	762
Binary Insurance Variables	
Private Insurance	973
Public Insurance (i.e., Medicaid, Tricare, CHIP)	1,925
More than one type	69
Self-pay/No insurance	58
Binary Child Age Variables	
Infant (0 - 12 months)	1,337
Toddler (13 - 35 months)	851
Preschool (36 – 59 months)	482
Children (60 – 71 months)	296

Kruskal-Wallis Tests

We used Kruskal-Wallis tests to evaluate differences across subgroups with more than two categories. These non-parametric tests allowed us to assess overall group differences without assuming a normal distribution of responses. The following multi-category variables were included:

1. *Insurance type* (public, private, more than one insurance, no insurance)
2. *Race/ethnicity* (seven categories, including multiracial and Hispanic/Latino)
3. *Child age group* (infants, toddlers, preschoolers, and children)

Results from the Kruskal–Wallis tests were used to identify where statistically significant differences occurred and to inform the selection of covariates for subsequent regression analyses.

Ordinal Logistic Regression

Finally, we conducted a series of ordinal logistic regression models to examine the effect of demographic characteristics while adjusting for potential confounding variables. This method is appropriate for ordinal survey responses and accounts for the ordered nature of Likert-scale items. Each model included:

1. *Dependent variables*: Individual CONNECT survey item scores (ordinal)
2. *Independent variables*: Race/ethnicity, insurance type, child age group
3. *Control variables*: Frequency of visits with the same provider and whether the caregiver typically brings the child to appointments
4. *Estimation method*: Robust standard errors to account for potential heteroskedasticity.

Results are presented as odds ratios (ORs). OR > 1.0 indicates greater odds of providing a higher (more positive) rating, and OR < 1.0 indicates lower odds of providing a higher rating. This multivariable approach allowed for a more nuanced understanding of how demographic factors relate to perceived relational quality, beyond what is captured by simple group comparisons.

Results by Subgroup

Race/Ethnicity

Wilcoxon Rank-Sum Test

The results of rank-sum analyses indicate that caregivers gave high ratings, reflecting generally positive impressions of their child's provider. However, closer examination shows subtle but meaningful differences (see Tables 12 – 15 in Appendix B).

Rank-Sum tests demonstrated that *White* caregivers consistently reported higher rating**s** than other groups on the following items: (1) *Overall average score* (M = 3.950 vs. 3.929, p = .013), (2) *My child's doctor made a connection with me and my child* (M = 3.936 vs. 3.906, p = .012), (3) *My child's doctor made me feel valued and respected* (M = 3.952 vs. 3.92, p = .005), (4) *My child's doctor made me feel safe and comfortable* (M = 3.956 vs. 3.929, p = .009), (5) *My child's doctor did not judge me* (M = 3.954 vs. 3.927, p = .014), and (6) *My child's doctor was kind and showed they cared* (M = 3.956 vs. 3.930, p = .024).

Black/African American caregivers expressed slightly lower ratings on the following: (1) *My child's doctor made a connection with me and my child* (M = 3.888 vs. 3.926, p = .041), (2) *My*

child's doctor did not judge me (M = 3.909 vs. 3.946, p = .022), and (3) *My child's doctor was kind and showed they cared* (M = 3.907 vs. 3.946, p = .005). Similarly, both *Multiracial* (M = 3.729 vs. 3.892, p = .005) and *Native American* caregivers (M = 3.785 vs. 3.892; Z = 2.325, p = .020) provided lower ratings on the CAHPS item *How often did your child's personal doctor explain things about your child's health in a way that was easy to understand.*

Kruskal-Wallis Test

The results of Kruskal-Wallis tests found statistically significant differences in item performance only on Press Ganey and CAHPS questions. On the Press Ganey question, *What is the likelihood of recommending this care provider to others?* ($\chi^2(6) = 23.325$, p = .001): (1) *White* caregivers were significantly more likely to recommend the provider than *Black* (p < .001), *Asian* (p = .002), *Hispanic* (p = .037), and *Multiracial* caregivers (p = .002), (2) *Black* caregivers were more likely to recommend than *Hispanic* caregivers (p = .041), (3) *Hispanic* caregivers were more likely to recommend than *Asian* (p = .027) and *Multiracial* caregivers (p = .024), and (4) *American Indian/Alaska Native* caregivers reported higher likelihood to recommend scores than *Asian* (p = .045) and *Multiracial* caregivers (p = .035) caregivers.

Across racial/ethnic groups, the CAHPS question, *How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?* ($\chi^2(6) = 23.02$, p = .001) demonstrated the following differences: (1) *White* caregivers reported significantly greater scores than *Black* (p = .003), *Asian* (p = .006), *Hispanic* (p = .030), and *Multiracial* caregivers (p < .001). (2) *Native Hawaiian/Other Pacific Islander* caregivers reported higher scores than *Black* (p = .024), *Asian* (p = .012), (3) *Hispanic* caregivers (p = .043), and *Multiracial* caregivers (p = .003), and (3) *Black* (p = .031), *Hispanic* (p = .008), and *American Indian/Alaska Native* (p = .017) caregivers reported higher scores than *Multiracial* caregivers.

Ordinal Logistic Regression

Ordinal logistic regression analyses examined how perceptions of the provider relationship and overall satisfaction varied across racial and ethnic groups, controlling for covariates. Table 16 (Appendix B) summarizes adjusted odds ratios (OR) and significance levels. The results indicate meaningful disparities in caregiver-reported experiences on the following items:

1. *My child's doctor made me feel valued and respected.*

- *Native American/Alaska Native* caregivers have 55% lower odds of feeling valued and respected by their child's doctor (OR = 0.45, *p* = .040).
- *White* caregivers had 52% higher odds of reporting feeling valued and respected (OR = 1.52, *p* = .048).

2. *My child's doctor was kind and showed they cared.*
    - *Black/African American* caregivers had 40% lower odds of perceiving their child's doctor as kind and caring (OR = 0.60, *p* = .030).
3. *My child's doctor talked about my child's health in the way I understood.*
    - *White* caregivers had 39% higher odds of reporting that their doctor explained things clearly (OR = 1.39, *p* = .022).
    - *Asian* (OR = 0.53, *p* = .020) and *Multiracial* (OR = 0.50, *p* = .026) caregivers had lower odds of reporting clear communication (47% and 50%, respectively).
4. *What is the likelihood of you recommending this care provider to others?* (Press Ganey)
    - *Black/African American* (OR = 0.74, *p* = .042), *Asian* (OR = 0.54, *p* = .009), and *Multiracial* (OR = 0.56, *p* = .035) caregivers had significantly lower odds of recommending their provider to others (26%, 46%, and 44%, respectively).
    - *White* caregivers had 55% higher odds of reporting that they would recommend their provider (OR = 1.55, *p* = .001).

Insurance Type

Wilcoxon Rank Sum Test

Across all survey items listed (Table 17 in Appendix C) *caregivers with public insurance* reported slightly lower mean ratings than those with other insurance types (private, none): (1) *Overall average score* (M = 3.930 vs. 3.953, *p* = .026), (2) *My child's doctor made me feel valued and respected* (M = 3.924 vs. 3.954, *p* = .01), (3) *My child's doctor made a connection with me and my child* (M = 3.910 vs. 3.942, *p* = .003), (4) *My child's doctor made me feel safe and comfortable* (M = 3.932 vs. 3.957, *p* = .027), (5) *My child's doctor did not judge me* (M = 3.930 vs. 3.955, *p* = .047), (6) *My child's doctor was kind and showed they cared* (M = 3.931 vs. 3.960, *p* = .011), (7) *My child's doctor talked about my child's health in the way I understood* (M = 3.937 vs. 3.963, *p* = .007), (8) *My child's doctor addressed my questions and concerns* (M = 3.936 vs. 3.957, *p* = .025), (9) *My child's*

doctor gave me the information on how to get in touch with them (M = 3.924 vs. 3.943, p = .046), and (10) *The doctor's office staff were respectful and professional* (M = 3.935 vs. 3.958, p = .036).

Similarly, Table 18 (Appendix C) shows that *caregivers with private insurance* gave significantly higher average ratings than those with other coverage types across key items: (1) *Overall average score* (M = 3.951 vs. 3.769, p = .039), (2) *My child's doctor made me feel valued and respected* (M = 3.951 vs. 3.926, p = .022), (3) *My child's doctor made a connection with me and my child* (M = 3.941 vs. 3.908, p = .003), (4) *My child's doctor made me feel safe and comfortable* (M = 3.955 vs. 3.934, p = .050), (5) *My child's doctor was kind and showed they cared* (M = 3.958 vs. 3.933, p = .021) (6) *My child's doctor talked about my child's health in the way I understood* (M = 3.961 vs. 3.938, p = .013), (7) *My child's doctor addressed my questions and concerns* (M = 3.957 vs. 3.937, p = .025), (8) *My child's doctor gave me information on how to get in touch with them* (M = 3.942 vs. 3.925, p = .039), and (9)*The doctor's office staff were respectful and professional* (M = 3.957 vs. 3.936, p = .037).

Kruskal-Wallis Test

The results of Kruskal-Wallis tests found statistically significant differences in item performance on a few CONNECT survey items observed by insurance type: (1) *My child's doctor made a connection with me and my child* (χ^2 (3) = 10.49, p = .015), *private insurance > public insurance* (p = .001), (2) *My child's doctor talked about my child's health in the way I understood* (χ^2 (3) = 7.98, p = .046), *private insurance > public insurance* (p = .004), (3) *What is the likelihood of you recommending this care provider to others?* (Press Ganey) (χ^2 (3) = 8.41, p = .038), *private insurance > public insurance* (p = .003), (4) *How often did your child's personal doctor explain things about your child in a way you understood?* (CAHPS) (χ^2 (3) = 16.31, p = .001), *private insurance > public insurance* (p < .001) and *private insurance > no insurance* (p = .006), and (5) *How often did your child's personal doctor show respect for what you had to say?* (CAHPS) (χ^2 (3) = 8.94, p = .030) *private insurance* (p = .007) and *public insurance* (p = .021) > *no insurance*.

Ordinal Logistic Regression

Table 19 in Appendix C highlights statistically significant differences in caregiver-reported experiences based on insurance type, controlling for covariates. Results indicate meaningful disparities in how caregivers with public vs. private insurance perceive pediatric care.

1. *My child's doctor made me feel valued and respected.*
    - *Caregivers with public insurance* had 39% lower odds of feeling valued and respected (OR = 0.61, *p* = .038).
2. *My child's doctor made a connection with me and my child.*
    - *Caregivers with public insurance* had 41% lower odds of perceiving their doctor making a connection with them and their child (OR = 0.59, *p* = .011).
    - *Caregivers with private insurance* had 65% higher odds of perceiving their doctor making a connection with them and their child (OR = 1.65, *p* = .014).
3. *My child's doctor was kind and showed me they cared.*
    - *Caregivers with public insurance* had 44% lower odds of perceiving their doctor as kind and caring (OR = 0.56, *p* = .030).
4. *My child's doctor talked about my child's health in the way I understood.*
    - *Caregivers with public insurance* had 52% lower odds of reporting clear communication (OR = 0.48, p = .008).
    - *Caregivers with private insurance* had 96% higher odds of reporting clear communication by their doctor (OR = 1.96, *p* = .016).
5. *My child's doctor addressed my questions and concerns.*
    - *Caregivers with public insurance* had 41% lower odds of reporting their doctor answered their questions and concerns (OR = 0.59, *p* = .047).
    - *Caregivers with private insurance* had 70% higher odds of reporting their doctor answered their questions and concerns (OR = 1.70, *p* = .049).
6. *My child's doctor gave me information on how to get in touch with them.*
    - *Caregivers with public insurance* had 35% lower odds of reporting their doctor gave them the information on how to get in touch with them (OR = 0.65, *p* = .048).
    - *Caregivers with private insurance* had 61% higher odds of reporting their provider gave them information on how to get in touch with them (OR = 1.61, *p* = .033).
7. *How often did your child's personal doctor explain things about your child in a way you understood?* (CAHPS)
    - *Caregivers with public insurance* had 35% lower odds of reporting their doctor explained things well (OR = 0.65, *p* = .007).

- *Caregivers with private insurance* had 69% higher odds of reporting clear communication (OR = 1.69, p = .002).

8. *What is the likelihood of you recommending this care provider to others?* (Press Ganey)
    - *Caregivers with public insurance* had 26% lower odds of recommending the care provider to others (OR = 0.74, p = .034).
    - *Caregivers with private insurance* had 34% higher odds of recommending the care provider to others (OR = 1.34, p = .039).

These findings suggest that caregivers with private insurance perceive their interactions with pediatric providers more favorably than those with public or no insurance, pointing to systemic differences in how families experience care based on coverage type.

Child Age

Wilcoxon Rank Sum Test

Caregivers of toddlers consistently reported more positive ratings than other age groups (Table 20, Appendix D). Specific findings include: (1) *My child's doctor made me feel safe and comfortable* (M = 3.966 vs. 3.930, p = .007), (2) *My child's doctor talked about my child's health in the way I understood* (M = 3.969 vs. 3.937, p = .014), (3) *My child's doctor addressed my questions and concerns* (M = 3.969 vs. 3.933, p = .005), (4) *My child's doctor gave me full attention* (M = 3.964 vs. 3.935, p = .049), and (5) *My child's doctor did not seem rushed during the visit* (M = 3.954 vs. 3.923, p = .039).

In contrast, *caregivers of preschoolers* (Table 21, Appendix D) provided somewhat lower ratings than other age groups. While differences were relatively small, each was statistically significant, indicating that *preschool caregivers* felt less positive about the pediatric visit experience than other age groups on the following items: (1) *My child's doctor made me feel safe and comfortable* (M = 3.919 vs. 3.945, p = .029), (2) *My child's doctor was kind and showed they cared* (M = 3.921 vs. 3.945, p = .020), (3) *My child's doctor talked about my child's health in the way I understood* (M = 3.925 vs. 3.950, p = .025), (4) *My child's doctor addressed my questions and concerns* (M = 3.921 vs. 3.948, p = .017), (5) *My child's doctor gave me full attention* (M = 3.919 vs. 3.946, p = .044), (6) *My child's doctor gave me information on how to get in touch with them* (M = 3.907 vs. 3.935, p = .016), (7)*The doctor's office staff were respectful and professional* (M = 3.921 vs. 3.947, p = .019),

and (8) *How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?* (CAHPS) (M = 3.833 vs. 3.896, p = .006).

Only one item reached statistical significance among *caregivers of children* ages five and older. Compared to other groups, these caregivers gave slightly lower ratings on the question *My child's doctor addressed my questions and concerns* (M = 3.892 vs. 3.949, p = .022) (Table 22, Appendix D).

Kruskal-Wallis Test

The results of Kruskal-Wallis tests found statistically significant differences in item performance on a few CONNECT survey items observed by child age: (1) *My child's doctor made me feel valued* and respected ($\chi^2(3) = 7.84$, p = .049), *preschoolers < infants* (p = .049) *and toddlers* (p = .014), *and children < infants* (p = .044) *and toddlers* (p = .015). (2) *My child's doctor made me feel safe and comfortable* ($\chi^2(3) = 12.38$, p = .006), *preschoolers < infants* (p = .026), *and toddlers* (p = .001), and *children < toddlers* (p = .017). (3) *My child's doctor was kind and showed they cared* ($\chi^2(3) = 9.93$, p = .019), *preschoolers < infants* (p = .015) *and toddlers* (p = .004), *and children < toddlers* (p = .021). (4) *My child's doctor gave me full attention* ($\chi^2(3) = 9.05$, p = .029), *preschoolers < infants* (p = .021), *and toddlers* (p = .012), *and children < infants* (p = .025) *and toddlers* (p = .014). (5) *My child's doctor worked with me in making decisions about my child's health* ($\chi^2(3) = 8.00$, p = .046), *preschoolers < infants* (p = .041), *and toddlers* (p = .024), *and children < infants* (p = .022) *and toddlers* (p = .013). (6) *My child's doctor talked about my child's health in the way I understood* ($\chi^2(3) = 10.78$, p = .013), *preschoolers < infants* (p = .022) *and toddlers* (p = .002), *and children < toddlers* (p = .010). (7) *My child's doctor addressed my questions and concerns* ($\chi^2(3) = 15.46$, p = .002), *preschoolers < infants* (p = .018) *and toddlers* (p = .001), *and children < infants* (p = .015) *and toddlers* (p = .001). (8) *My child's doctor gave me information on how to get in touch with them* ($\chi^2(3) = 8.25$, p = .041), *preschoolers < infants* (p = .009) *and toddlers* (p = .007). (9) *The doctor's office staff were respectful and professional* ($\chi^2(3) = 9.99$, p = .019), *preschoolers < infants* (p = .015) *and toddlers* (p = .003), *and children < toddlers* (p = .017).

Differences were also found on the CAHPS question *How often did your child's personal doctor explain things in a way that was easy to understand?* (CAHPS) ($\chi^2(3) = 8.33$, p = .040) with *preschool caregivers < infants* (p = .007) *and toddlers* (p = .003).

Ordinal Logistic Regression

Results from ordinal logistic regression (Table 23, Appendix D) further support these patterns. After adjusting for covariates, the following patterns emerged:

1. *My child's doctor made me feel safe and comfortable*.
    - *Caregivers of toddlers* had 86% higher odds of reporting feeling safe and comfortable (OR = 1.86, p = .022).
2. *My child's doctor talked about my child's health in the way I understood*.
    - *Caregivers of toddlers* had 78% higher odds of reporting clear communication (OR = 1.78, p = .047).
3. *My child's doctor addressed my questions and concerns*.
    - *Caregivers of toddlers* had 97% higher odds of reporting their doctor addressed their questions and concerns (OR = 1.97, p = .020).
4. *How often did your child's personal doctor explain things about your child's health in a way that was easy to understand (CAHPS)?*
    - *Caregivers of preschoolers* had 33% lower odds of reporting their doctor explained things clearly (OR = 0.67, p = .035).

These findings suggest that child age may meaningfully shape how caregivers perceive pediatric care interactions. Toddlers' caregivers experience the most positive visits, and preschool caregivers report comparatively less favorable interactions.

Summary

Statistically significant differences in caregiver-reported experiences emerged across all three demographic domains:

- **Race/Ethnicity:** White caregivers consistently reported more positive experiences, while Black, Multiracial, Asian, and Native American caregivers were more likely to report lower ratings.
- **Insurance Type:** Caregivers with public insurance reported less favorable experiences than those with private insurance.
- **Child Age:** Toddler caregivers gave the most positive ratings, while preschool caregivers reported lower scores, suggesting developmental stage may influence how visits are experienced and perceived.

Discussion and Implications

Interpretation of Key Findings

The CONNECT survey was developed to address the absence of validated, parent-informed tools to assess the quality of the parent/caregiver-clinician relationship. Results from this validation study provide strong evidence of the instrument's reliability, validity, and practical utility.

- **Reliability**: The survey demonstrated excellent internal consistency (Cronbach's α = 0.99), with inter-item correlations ranging from r = .63 to .89. These findings confirm that the items cohesively measure a single underlying construct—relational quality. However, the strength of these correlations suggests potential redundancy across some items, which could be considered in future refinements.
- **Validity**: Exploratory factor analysis identified one dominant factor, accounting for over 98% of the variance, supporting the interpretation of the survey as a unidimensional tool. Although items were initially categorized under three parent-identified themes, the factor structure suggests that caregivers experience these relational elements as highly interconnected.
- **Convergent Validity**: Acceptable correlations with items from the CAHPS and Press Ganey surveys support the convergent validity of the CONNECT survey. While correlation values

were modest, the direction and pattern of associations were consistent with expectations, particularly considering differences in item format and response scale.

- **Measurement Invariance**: Analyses indicate significant subgroup differences across race/ethnicity, insurance status, and child age. These findings suggest that the CONNECT survey is sensitive to disparities in caregiver/provider relationships and make it a valuable tool for quality monitoring and equity-oriented improvement efforts. For example:
    - White and privately insured caregivers reported more favorable experiences than those from historically marginalized groups or with public insurance.
    - Caregivers of toddlers reported the most positive experiences, while caregivers of preschoolers and older children reported less favorable perceptions in several areas.

Together, these findings provide evidence that the CONNECT survey captures meaningful variation in caregiver-reported experiences and can help identify relational strengths and gaps across diverse populations.

Comparison to Other Measures

The CONNECT survey differs from existing tools in several ways:

- **Parent-Centered Development**: Unlike many existing instruments developed by researchers or clinicians, the CONNECT survey was co-designed with parents and caregivers, capturing parent concepts and language with respect to their relationship with the primary care clinician.
- **Pediatric-Specific and Visit-Specific**: While tools like CAHPS focus broadly on system-level experiences, the CONNECT survey captures visit-level relational quality specific to pediatric care.
- **Brevity and Accessibility**: The 13-item format, written in parent-friendly language, allows for rapid post-visit administration without compromising psychometric rigor.

The high overall scores in this study resemble patterns seen in Press Ganey and CAHPS data, which also often exhibit ceiling effects. However, the CONNECT survey may offer a more sensitive lens into caregiver relationships by explicitly focusing on respect, connection, and communication, constructs that are foundational to early relational health.

Strengths and Limitations

Strengths

- The survey was co-designed with a racially and geographically diverse group of caregivers.
- It demonstrates strong internal consistency and construct validity.
- The tool reveals meaningful subgroup differences, underscoring its potential as an equity-focused measurement instrument.
- The instrument can be integrated easily into routine care workflows, especially immediately after well-child visits.

Limitations

- The high mean scores and clustering of responses at the top of the scale may indicate ceiling effects, potentially limiting sensitivity in differentiating providers or care settings.
- The strong inter-item correlations suggest some content overlap among items.
- The high test-retest reliability values also suggest redundancy among items.
- Sites participating in the validation study received high Reach Out and Read Quality Site Classification scores (13.41% yellow-classified, 86.59% green-classified). This may have influenced high top-box/ceiling effects and impacted statistical findings.

Implications

The findings imply that relationship quality significantly varies across demographic groups. This highlights opportunities to improve parent-clinician relationships involving communication and trust, especially for families with public insurance and certain racial/ethnic groups. Pediatric practices can use this measurement tool to routinely measure and improve relational quality, potentially improving overall care quality and child health outcomes.

Conclusion and Recommendations

Summary of Validation Outcomes

Findings indicate that the CONNECT survey is a reliable and valid parent-informed measure of relationship quality between parents/caregivers and pediatric clinicians. The survey:

- Demonstrates **excellent internal consistency** and a strong **unidimensional factor structure**, confirming that it captures a cohesive construct of relational quality.
- Exhibits **convergent validity** through significant associations with established patient experience measures.
- Identifies **meaningful subgroup differences**, confirming the tool's sensitivity to variations across race/ethnicity, insurance type, and child age group.
- Was **developed in direct partnership with parents and caregivers**, ensuring it reflects the values, language, and lived experiences of families in pediatric care settings.

While scores were consistently high across items, response skewness and limited variability suggest that further refinements may enhance the survey's sensitivity to more subtle differences in care experiences.

Future Goals

To strengthen the utility and reach of the CONNECT survey, we plan to carry out future refinement, implementation, and research:

1. *Refine the Response Scale*
    - Transition from a 4-point to a 5-point or 7-point response scale to reduce ceiling effects and allow for greater differentiation among high-performing clinicians or clinics.
2. *Address Item Redundancy*
    - Based on the high inter-item correlations observed, review and potentially consolidate items with overlapping content. Reducing redundancy can shorten the survey without compromising measurement precision.
3. *Enhance Visual Design and Flow*
    - Revise formatting and layout to minimize cognitive breaks between survey sections (e.g., consistent color scheme, unified scale language), particularly when embedding external items such as CAHPS or Press Ganey measures.

4. *Conduct Additional Validity and Reliability Testing*
    - Evaluate the survey's performance in more varied clinical environments, including non-Reach Out and Read sites and clinics with lower baseline relational scores.
5. *Use the Tool for Equity-Oriented Improvement*
    - To monitor disparities in care experience, regularly stratify CONNECT survey data by key demographic variables (i.e., insurance type, race/ethnicity, and age group).
    - Integrate the tool into ongoing quality improvement (QI) initiatives to advance equity in pediatric care.

References

Agency for Healthcare Research and Quality. (2023). *CAHPS® surveys and guidance*. U.S. Department of Health & Human Services. Retrieved May 12, 2025, from https://www.ahrq.gov/cahps/surveys-guidance/index.html

Alchemer. (n.d.). *How survey question order impacts results*. Retrieved May 12, 2025, from https://www.alchemer.com/resources/blog/survey-question-order/

American Academy of Family Physicians. (2022). How to improve patient experience in your practice. *Family Practice Management, 29*(2), 27–32. https://www.aafp.org/pubs/fpm/issues/2022/0300/p27.html

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs, 27*(3), 759–769. https://doi.org/10.1377/hlthaff.27.3.759

Bethell, C. D., Garner, A. S., Gombojav, N., Blackwell, C., Heller, L., & Mendelson, T. (2022). Social and relational health risks and common mental health problems among US children: The mitigating role of family resilience and connection to promote positive socioemotional and school-related outcomes. *Child and Adolescent Psychiatric Clinics of North America, 31*(1), 45–70. https://doi.org/10.1016/j.chc.2021.08.001

Burton, H., & Navsaria, D. (2019). Evaluating the effect of Reach Out and Read on clinic values, attitudes, and knowledge. *WMJ: Official Publication of the State Medical Society of Wisconsin, 118*(4), 177–181.

Center for the Study of Social Policy. (2020) My Baby, My Doctor & Me: Family Voice in Early Relational Health Pilot Study. https://cssp.org/resource/my-baby-my-doctor-me-family-voice-in-early-relational-health-pilot-study/

Committee on Hospital Care & Institute for Patient- and Family-Centered Care. (2012). Patient andfamily-centered care and the pediatrician's role. *Pediatrics, 129*(2), 394–404. https://doi.org/10.1542/peds.2011-3084

Cordoba, D. J., Levin, E. R., Ramachandran, U., Lima, D., Shearman, N., Willis, D., ... & Jimenez, M. E. (2024). Parents' Perspectives on Early Relational Health: A Qualitative Study. *Journal of Developmental & Behavioral Pediatrics, 45*(4), e334-e340. https://doi.org/10.1097/DBP.0000000000001301

Garner, A., & Yogman, M. (2021). Preventing childhood toxic stress: Partnering with families and communities to promote relational health. *Pediatrics, 148*(2), Article e2021052582. https://doi.org/10.1542/peds.2021-052582

Gotschall, J. W., Fitzsimmons, R., Shin, D. B., & Takeshita, J. (2024). Race, ethnicity, and other patient and clinical encounter characteristics associated with patient experiences of access to care. *Journal of Patient Experience*, *11*, 23743735241241178. https://doi.org/10.1177/23743735241241178

Haviland, M. G., Morales, L. S., Dial, T. H., & Pincus, H. A. (2005). Race/ethnicity, socioeconomic status, and satisfaction with health care. American Journal of Medical Quality, 20(4), 195–203. https://doi.org/10.1177/1062860605275754

IMPACT Measures Tool: Scoring system evidence guide. (2002). Institute for Child Success. https://www.instituteforchildsuccess.org/our-initiatives/initiative/impact-measures-tool/

Johnson, R. L., Roter, D., Powe, N. R., & Cooper, L. A. (2004). Patient race/ethnicity and quality of patient-physician communication during medical visits. *American Journal of Public Health (1971)*, *94*(12), 2084–2090. https://doi.org/10.2105/AJPH.94.12.2084

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Murray-García, J. L., Selby, J. V., Schmittdiel, J., Grumbach, K., & Quesenberry, C. P. (2000). Racial and ethnic differences in a patient survey: Patients' values, ratings, and reports regarding physician primary care performance in a large health maintenance organization. *Medical Care*, *38*(3), 300–310. https://doi.org/10.1097/00005650-200003000-00007

North, F., & Tulledge-Scheitel, S. M. (2019). Patient satisfaction with providers: Do patient surveys give enough information to help providers improve specific behaviors? *Health Services Research and Managerial Epidemiology, 6*, 2333392819885284. https://doi.org/10.1177/2333392819885284

Petts, R. A., & Shahidullah, J. D. (2020). Engagement interventions delivered in primary care to improve off-site pediatric mental health service initiation: A systematic review. *Families, Systems, & Health, 38*(4), 456–468. https://doi.org/10.1037/fsh0000521

Pew Research Center. (2020, January 29). *Writing survey questions*. Retrieved May 12, 2025, from https://www.pewresearch.org/writing-survey-questions/

Press Ganey. (n.d.). *Patient experience survey*. Press Ganey Associates LLC. Retrieved May 12, 2025, from https://www.pressganey.com/

Reach Out and Read. (2024). *Reach Out and Read: Site coordinator manual, 2023-2024*. Reach Out and Read. Retrieved May 12, 2025, from https://reachoutandread.org/wp-content/uploads/2024/10/2024-Site-Coordinator-Manual.pdf

Turpin, R. E., Akré, E.-R. L., Williams, N. D., Boekeloo, B. O., & Fish, J. N. (2021). Differences in health care access and satisfaction ccross intersections of race/ethnicity and sexual Identity. *Academic Medicine*, *96*(11), 1592–1597. https://doi.org/10.1097/ACM.0000000000004243

Vaughn, N. & Snively, S. (2024). Relias. Maximizing healthcare reimbursement through higher patient satisfaction scores. *Relias Blog*. Retrieved May 12, 2025, from https://www.relias.com/blog/healthcare-reimbursement-patient-satisfaction-scores

Willis, D. W., & Eddy, J. M. (2022). Early relational health: Innovations in child health for promotion, screening, and research. *Infant Mental Health Journal, 43*(3), 361–372. https://doi.org/10.1002/imhj.21984

Appendix A: Item Correlations Between CONNECT Survey Items

Table 10: Spearman Correlations Between CONNECT Survey Items

	CONNECT item1	CONNECT item 2	CONNECT item 3	CONNECT item 4	CONNECT item 5	CONNECT item 6	CONNECT item 7	CONNECT item 8	CONNECT item 9	CONNECT item 10	CONNECT item 11	CONNECT item 12	CONNECT item 13
CONNECT item 1													
CONNECT item 2	0.73												
CONNECT item 3	0.84	0.74											
CONNECT item 4	0.83	0.71	0.86										
CONNECT item 5	0.80	0.71	0.86	0.84									
CONNECT item 6	0.78	0.67	0.81	0.82	0.84								
CONNECT item 7	0.72	0.69	0.74	0.74	0.79	0.78							
CONNECT item 8	0.81	0.71	0.82	0.85	0.82	0.83	0.78						
CONNECT item 9	0.81	0.70	0.80	0.80	0.80	0.80	0.73	0.84					
CONNECT item 10	0.77	0.66	0.81	0.79	0.80	0.81	0.75	0.82	0.84				
CONNECT item 11	0.77	0.66	0.81	0.77	0.79	0.80	0.72	0.79	0.84	0.89			
CONNECT item 12	0.69	0.68	0.72	0.72	0.73	0.71	0.72	0.74	0.81	0.78	0.78		
CONNECT item 13	0.74	0.63	0.77	0.75	0.79	0.77	0.70	0.77	0.80	0.83	0.82	0.77	

Table 11: Polychoric Correlations Between CONNECT Survey Items

	CONNECT item1	CONNECT item 2	CONNECT item 3	CONNECT item 4	CONNECT item 5	CONNECT item 6	CONNECT item 7	CONNECT item 8	CONNECT item 9	CONNECT item 10	CONNECT item 11	CONNECT item 12	CONNECT item 13
CONNECT item 1													
CONNECT item 2	0.97												
CONNECT item 3	0.99	0.98											
CONNECT item 4	0.99	0.97	0.99										
CONNECT item 5	0.99	0.97	0.99	0.99									
CONNECT item 6	0.98	0.96	0.99	0.99	0.99								
CONNECT item 7	0.97	0.96	0.97	0.97	0.98	0.98							
CONNECT item 8	0.98	0.96	0.99	0.98	0.99	0.98	0.98						
CONNECT item 9	0.098	0.97	0.99	0.98	0.98	0.99	0.97	0.99					
CONNECT item 10	0.98	0.97	0.99	0.99	0.99	0.99	0.98	0.99	0.99				
CONNECT item 11	0.98	0.96	0.99	0.98	0.98	0.98	0.97	0.98	0.99	0.99			
CONNECT item12	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.97	0.98	0.98	0.96		
CONNECT item13	0.97	0.95	0.98	0.98	0.98	0.98	0.97	0.98	0.99	0.99	0.99	0.98	

Appendix B: Significant Results by Items for Race/Ethnicity

Table 12: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for White Group

Item	White			All other races			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
Average Score	3.950	0.318	4	3.929	0.376	4	2.490	0.013
My child's doctor made a connection with me and my child.	3.936	0.354	4	3.906	0.417	4	2.521	0.012
My child's doctor did not judge me.	3.954	0.323	4	3.927	0.400	4	2.458	0.014
My child's doctor was kind and showed me they cared.	3.956	0.321	4	3.930	0.396	4	2.250	0.024
My child's doctor made me feel valued and respected.	3.952	0.326	4	3.921	0.406	4	2.827	0.005
My child's doctor made me feel safe and comfortable.	3.956	0.321	4	3.929	0.391	4	2.632	0.009

Note. SD = Standard Deviation.

Table 13: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Black/African American Group

Item	Black/African American			All other races			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
My child's doctor made a connection with me and my child.	3.888	0.465	4	3.926	0.371	4	2.043	0.041
My child's doctor did not judge me.	3.909	0.450	4	3.946	0.346	4	2.289	0.022
My child's doctor was kind and showed me they cared.	3.907	0.460	4	3.946	0.3346	4	2.787	0.005

Note. SD = Standard Deviation.

Table 14: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Multi-Racial American Group

Item	Multi-racial			All other races			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
Average Score	3.924	0.414	4	3.938	0.350	4	1.558	0.119
How often did your child's personal doctor explain things about your child in a way you could understand?	3.729	0.721	4	3.892	0.402	4	2.800	0.005

Note. SD = Standard Deviation.

Table 15: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for American Indian/Alaska Native Group

Item	American Indian/Alaska Native			All other races			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
How often did your child's personal doctor explain things about your child in a way you could understand?	3.785	0.626	4	3.892	0.403	4	2.325	0.020

Note. SD = Standard Deviation.

Table 16: Ordinal Logistic Regression Results by Significant Items and Race/Ethnicity

Item	Native American/Alaska Native (OR, *p*)	Black / African American (OR, *p*)	White (OR, p)	Asian (OR, *p*)	Multiracial (OR, *p*)
My child's doctor made me feel valued and respected.	0.454, $p = .040$	—	1.522, $p = .048$	—	—
My child's doctor was kind and showed me they cared.	—	0.597, $p = .030$	—	—	—
How often did your child's personal doctor explain things in an easy way?	—	—	1.392, $p = .022$	0.528, $p = .020$	0.503, $p = .026$

Item	Native American/Alaska Native (OR, *p*)	Black / African American (OR, *p*)	White (OR, p)	Asian (OR, *p*)	Multiracial (OR, *p*)
What is the likelihood of you recommending this care provider to others?	—	0.741, *p* = .042	1.546, *p* = .001	0.544, *p* = .009	0.563, *p* = .035

Notes: (a) Cells show Odds Ratios (OR) with their corresponding *p*-values; (b) A dash (–) indicates that the race/ethnicity comparison was not significant; (c) All odds ratios are interpreted relative to the reference group (e.g., all others), adjusting for covariates; (d) OR > 1.0 suggests higher odds of reporting a higher score; OR < 1.0 suggests lower odds of reporting a higher score.

Appendix C: Significant Results by Items for Type of Insurance

Table 17: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Individuals with Public Insurance

Item	Public Insurance			All others			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
Average Score	3.930	0.373	4	3.953	0.308	4	2.223	0.0260
My child's doctor made me feel valued and respected.	3.924	0.400	4	3.954	0.319	4	2.584	0.01
My child's doctor made a connection with me and my child.	3.910	0.417	4	3.942	0.338	4	3.016	0.003
My child's doctor made me feel safe and comfortable.	3.932	0.386	4	3.957	0.315	4	2.213	0.027
My child's doctor did not judge me.	3.930	0.393	4	3.955	0.318	4	1.988	0.047
My child's doctor was kind and showed me they cared.	3.931	0.392	4	3.960	0.311	4	2.536	0.011
My child's doctor talked about my child's health in the way I understood.	3.937	0.376	4	3.963	0.306	4	2.678	0.007
My child's doctor addressed my questions and concerns.	3.936	0.376	4	3.957	0.317	4	2.245	0.025
My child's doctor gave me information on how to get in touch with them.	3.924	0.396	4	3.943	0.356	4	1.998	0.046
The doctor's office staff were respectful and professional.	3.935	0.378	4	3.958	0.314	4	2.096	0.036

Note. SD = Standard Deviation.

Table 18: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Individuals with Private Insurance

Item	Private Insurance			All Others			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
Average Score	3.951	0.317	4	3.769	0.368	4	-2.063	0.039
My child's doctor made me feel valued and respected.	3.951	0.328	4	3.926	0.394	4	-2.286	0.022
My child's doctor made a connection with me and my child.	3.941	0.345	4	3.908	0.411	4	-2.936	0.003
My child's doctor made me feel safe and comfortable.	3.955	0.324	4	3.934	0.380	4	-1.963	0.050
My child's doctor was kind and showed me they cared.	3.958	0.320	4	3.933	0.386	4	-2.310	0.021
My child's doctor talked about my child's health in the way I understood.	3.961	0.315	4	3.938	0.371	4	-2.491	0.013
My child's doctor addressed my questions and concerns.	3.957	0.324	4	3.937	0.371	4	-2.243	0.025

Item	Private Insurance			All Others			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
My child's doctor gave me information on how to get in touch with them.	3.942	0.364	4	3.925	0.391	4	-2.061	0.039
The doctor's office staff were respectful and professional.	3.957	0.321	4	3.936	0.373	4	-2.088	0.037

Note. SD = Standard Deviation.

Table 19: Ordinal Logistic Regression Results by Significant Items and Type of Insurance

Item	Public Insurance (OR, *p*)	Private Insurance (OR, *p*)
My child's doctor made me feel valued and respected.	0.613, *p* = .038	–
My child's doctor made a connection with me and my child.	0.594, *p* = .011	1.653, *p* = .014
My child's doctor was kind and showed me they cared.	0.564, *p* = .030	–
My child's doctor talked about my child's health in the way I understood.	0.476, *p* = .008	1.956, *p* = .016
My child's doctor addressed my questions and concerns.	0.586, *p* = .047	1.699, *p* = .049
My child's doctor gave me information on how to get in touch with them.	0.647, *p* = .048	1.608, *p* = .033
How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?	0.647, *p* = .007	1.689, *p* = .002
What is the likelihood of you recommending this care provider to others?	0.740, *p* = .034	1.343, *p* = .039

Notes: (a) Cells show Odds Ratios (OR) with their corresponding *p*-values; (b) A dash (–) indicates that the type of insurance comparison was not significant; (c) All odds ratios are interpreted relative to the reference group (e.g., all others), adjusting for covariates; (d) OR > 1.0 suggests higher odds of reporting a higher score; OR < 1.0 suggests lower odds of reporting a higher score.

Appendix D: Significant Results by Items for Child Age

Table 20: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Toddler (13-35 months) Group

Item	Toddler			All other ages			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
My child's doctor made me safe and comfortable.	3.966	0.274	4	3.930	0.392	4	-2.685	0.007
My child's doctor talked about my child's health in the way I understood.	3.969	0.270	4	3.937	0.382	4	-2.457	0.014
My child's doctor addressed my questions and concerns.	3.969	0.270	4	3.933	0.385	4	-2.827	0.005
My child's doctor gave me full attention.	3.964	0.278	4	3.935	0.384	4	-1.971	0.049
My child's doctor did not seem rushed during the visit.	3.954	0.305	4	3.923	0.400	4	-2.060	0.039

Note. SD = Standard Deviation.

Table 21: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Preschoolers (36-59 months) Group

Item	Preschoolers			All other ages			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
My child's doctor made me safe and comfortable.	3.919	0.402	4	3.945	0.354	4	2.332	0.029
My child's doctor was kind and showed they cared.	3.921	0.400	4	3.945	0.358	4	2.334	0.020
My child's doctor talked about my child's health in the way I understood.	3.925	0.396	4	3.950	0.345	4	2.236	0.025
My child's doctor addressed my questions and concerns.	3.921	0.400	4	3.948	0.347	4	2.395	0.017
My child's doctor gave me full attention.	3.919	0.412	4	3.946	0.350	4	2.012	0.044
My child's doctor gave me information on how to get in touch with them.	3.907	0.420	4	3.935	0.375	4	2.420	0.016
The doctor's office staff were respectful and professional.	3.921	0.400	4	3.947	0.348	4	2.340	0.019
How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?	3.833	0.269	4	3.896	0.396	4	2.747	0.006

Note. SD = Standard Deviation.

Table 22: Wilcoxon Rank Sum Test Results by Average Scores and Statistically Significant Items for Children (60-71 months) Group

Item	Children			All other ages			Z	*p*
	Mean	SD	Median	Mean	SD	Median		
My child's doctor addressed my questions and concerns.	3.892	0.260	4	3.949	0.335	4	2.300	0.022

Note. SD = Standard Deviation.

Table 23: Ordinal Logistic Regression Results by Significant Items and Child Age

Item	Toddlers (OR, *p*)	Preschoolers (OR, *p*)
My child's doctor made me feel safe and comfortable.	1.860, $p = .022$	—
My child's doctor talked about my child's health in the way that I understood	1.780, $p = .047$	—
My child's doctor addressed my questions and concerns.	1.970, $p = .020$	—
How often did your child's personal doctor explain things about your child's health in a way that was easy to understand?	—	0.667, $p = .035$

Notes: (a) Cells show Odds Ratios (OR) with their corresponding *p*-values; (b) A dash (–) indicates that the race/ethnicity comparison was not significant; (c) All odds ratios are interpreted relative to the reference group (e.g., all others), adjusting for covariates; (d) OR > 1.0 suggests higher odds of reporting a higher score; OR < 1.0 suggests lower odds of reporting a higher score.