ARTICLE

The Home Visit Rating Scales: Revised, restructured, and revalidated

Lori A. Roggman[1] (iD) | Gina A. Cook[2] (iD) | Mark S. Innocenti[3] (iD) | Vonda K. Jump Norman[4] (iD) | Lisa K. Boyce[1] (iD) | Tasha L. Olson[1] (iD) | Katie Christiansen[1] (iD) | Carla A. Peterson[5] (iD)

[1]Human Development and Family Studies Department, Utah State University, Logan, Utah

[2]Department of Psychology and Child Development, California State University Stanislaus, Turlock, California

[3]Center for Persons with Disabilities, Utah State University, Logan, Utah

[4]Department of Sociology, Social Work, and Anthropology, Utah State University, Logan, Utah

[5]Department of Human Development and Family Studies, Iowa State University, Ames, Iowa

Correspondence
Lori A. Roggman, Human Development and Family Studies Department, Utah State University, Logan, UT 84322.
Email: Lori.Roggman@usu.edu

ABSTRACT

The Home Visit Rating Scales (HOVRS) were initially developed from field-based descriptions of successful home visits and are supported by home-visiting research in multiple disciplines. Four home-visiting practices scales include indicators of relationship building with families, responsiveness to family strengths, facilitation of parent–child interaction, and collaboration with parents. Three family engagement scales include indicators of parent–child interaction, parent engagement, and child engagement in the visit. The original version, the HOVRS-1, was validated using video and data from two Early Head Start home-visiting programs. Conceptual and structural changes for the HOVRS-3 were designed to improve readability, usability, and clarity. Newly trained observers used the HOVRS-3 to observe archived videos from the original measurement sample. The HOVRS-3 showed good interrater reliability, scale internal consistency, convergent validity, predictive validity, practical significance, and version stability. When the HOVRS-3 home-visit quality scores were higher, it was twice as likely for parenting scores to be average or better and for child language to be at age level or better at age 3 years, over and above parenting and child language at age 1 year. The HOVRS can guide observations of home-visit quality in infant–toddler and early childhood programs to improve home-visiting practices and family engagement.

KEYWORDS

family engagement, home-visiting practices, home-visit quality, Home Visit Rating Scales

RESUMEN

Las *Escalas de Evaluación de Visitas a Casa* inicialmente se desarrollaron a partir de las descripciones sobre el campo de exitosas visitas a casa y las mismas están apoyadas por la investigación sobre la visita a casa en múltiples disciplinas. Las escalas de prácticas de cuatro visitas a casa incluyen indicadores de relaciones establecidas con familias, sensibilidad hacia los puntos fuertes de la familia, el facilitar la interacción entre progenitor y niño, y la colaboración con los progenitores. Las escalas de participación de tres familias incluyen indicadores sobre la interacción entre progenitor y niño, la participación del progenitor, y la participación del niño en la visita. Se validó la versión original, *HOVRS-1*, usando un video e información

de dos programas de visitas a casa de Un Comienzo Temprano (Early Head Start). Se diseñaron cambios conceptuales y estructurales para *HOVRS-3* para mejorar la legibilidad, la disponibilidad para el uso y la claridad. Nuevos observadores entrenados usaron *HOVRS-3* para observar videos archivados acerca de las medidas de la muestra original – *HOVRS-3* mostró buena confiabilidad entre los evaluadores, consistencia interna de la escala, validez convergente, validez de predicción, importancia práctica y estabilidad de la versión. Cuando los puntajes de calidad de HOVRS-3 sobre la visita a casa fueron más altos, la inclinación de los puntajes de crianza de ser promedio o mejores fue dos veces mayor, así como también fue el lenguaje del niño de estar al nivel de la edad o mejor a los 3 años, muy por encima del lenguaje de crianza y del niño a la edad de un año. Las *Escalas de Evaluación de Visitas a Casa* pueden guiar las observaciones de la calidad de la visita a casa en programas para infantes y niños pequeñitos y en la temprana niñez para mejorar las prácticas de visitas a casa y la participación de la familia.

PALABRAS CLAVES

participación de la familia, prácticas de visitas a casa, calidad de la visita a casa, *Escalas de Evaluación de Visitas a Casa*

RÉSUMÉ

Les Echelles d'Evaluation de la Visite à Domicile ont été initialement développées à partir de descriptions sur le terrain de visites à domicile réussies et sont soutenues par les recherches sur les visites à domicile dans de multiples disciplines. Quatre échelles de pratique de visites à domicile incluent des indicateurs de développement de la relation avec les familles, la réaction aux forces familiales, la facilitation de l'interaction parent-enfant, et la collaboration avec les parents. Trois échelles d'engagement de la famille incluent des indicateurs d'interaction parent-enfant, d'engagement parental, et d'engagement de l'enfant durant la visite. La version originale, abrégée en anglais *HOVRS-1,* a été validée en utilisant des vidéos et des données de deux programmes de visites à domiciles américains dans le cadre du programme d'aide aux enfants défavorisés de Early Head Start. Les changements conceptuels et structurels du *HOVRS-3* ont été faits afin d'améliorer sa lisibilité, sa facilité d'utilisation et sa clarté. De nouveaux observateurs fraîchement formés ont utilisé les afin d'observer des vidéos mises en archive de l'échantillon original de mesure. Les *HOVRS-3* ont fait preuve d'une bonne fiabilité d'inter-évaluateur, d'une bonne cohérence interne à l'échelle, de validité convergente et de stabilité de version. Lorsque les scores de qualité de la visite à domicile HOVRS-3 étaient plus élevés il était deux fois plus probable que les scores de parentage soient moyens ou mieux et pour le langage de l'enfant qu'il soit au niveau de l'âge ou mieux à l'âge de 3 ans, bien au dessus du parentage et du langage de l'enfant à l'âge de 1 ans. Les *Echelles d'Evaluation de la Visite à Domicile* peuvent guider des observations de la qualité de la visite à domicile chez les nourrissons-petits-enfants et de programmes de la petite enfance afin d'améliorer les pratiques de visites à domicile et d'engagement de la famille.

MOTS CLÉS

engagement familial, pratiques de visite à domicile, qualité de la visite à domicile, *Echelles d'Evaluation de la Visite à Domicile*

ZUSAMMENFASSUNG

Die *Home Visit Rating Scales* wurden zunächst anhand feldbasierter Beschreibungen erfolgreicher Hausbesuche entwickelt und werden durch die Hausbesuchsforschung in mehreren Disziplinen unterstützt. Vier Skalen für die Hausbesuchspraktiken umfassen Indikatoren für den Beziehungsaufbau mit den Familien, die Responsivität auf familiäre Stärken, die Förderung der Eltern-Kind-Interaktion und die Zusammenarbeit mit den Eltern. Drei Skalen für das familiäre Engagement beinhalten Indikatoren für die Eltern-Kind-Interaktion, das Engagement der Eltern und das Engagement der Kinder während des Besuchs. Die Originalversion, *HOVRS-1*, wurde anhand von Videos und Daten aus zwei Early Head Start Hausbesuchsprogrammen validiert. Konzeptionelle und strukturelle Änderungen wurden für *HOVRS-3* vorgenommen, um die Lesbarkeit,

Benutzerfreundlichkeit und Übersichtlichkeit zu verbessern. Neu geschulte Beobachter nutzten *HOVRS-3* um archivierte Videos der ursprünglichen Stichprobe zu betrachten. *HOVRS-3* zeigte eine gute Interrater-Reliabilität, interne Konsistenz, konvergente Validität, prädiktive Validität, praktische Bedeutsamkeit und Versionsstabilität. Wenn die *HOVRS-3*-Werte zur Hausbesuchsqualität höher waren, war es doppelt so wahrscheinlich, dass die Ergebnisse der Elternkompetenz durchschnittlich oder besser waren und dass die Sprache des Kindes im Alter von 3 Jahren altersgerecht oder besser war, zusätzlich zur Elternkompetenz und Sprache des Kindes im Alter von 1 Jahr. Die *Home Visit Rating Scales* können Beobachtungen der Hausbesuchsqualität bei Säuglings- und Kleinkind-Programmen lenken, um die Hausbesuchspraktiken und das Engagement der Familien zu verbessern.

STICHWÖRTER

Engagement der Familien, Hausbesuchspraktiken, Hausbesuchsqualität, *Home Visit Rating Scales*

抄録

家庭訪問評価尺度は当初、家庭訪問の成功例を記述するというフィールドベースから発達し、複数分野での家庭訪問調査によりサポートされてきた。 家庭訪問実施尺度には、家族との関係構築、家族の持つ強みへの対応、親子相互作用の促進、および両親との協力の4つの指標が含まれている。

家族関与尺度には、訪問時の親子相互作用、親の関与、および子ども関与の3つの指標が含まれる。オリジナルバージョンの HOVRS-1 は、2つのEarly Head Start家庭訪問プログラムのビデオとデータを使用いて妥当性が検証された。 HOVRS-3 は、概念上および構造上読みやすさ、使いやすさ、および明快さを向上させるためにデザイン設計された。新たに訓練された観察者は HOVRS-3 を使用して元の測定サンプルからアーカイブされたビデオを観察した。 HOVRS-3 は、優れた評価者間信頼性、規模の内部一貫性、収束的妥当性、予測的妥当性、実用的意義、およびバージョン安定性を示した。 HOVRS-3 家庭訪問の質評価がより高かった場合、養育スコアは平均かそれ以上であり、3歳時点での子どもの言語が年齢レベルかそれ以上であった。1歳時点で親の養育と子どもの言語が年齢以上か年齢以上にずっと勝っているというように2倍の可能性があった。家庭訪問評価尺度は、乳幼児と乳幼児期のプログラムにおける家庭訪問の質に関する観察結果を導き、家庭訪問の実践と家族の関与を向上させることができる。

キーワード

家族の関与, 家庭訪問の実施, 家庭訪問の質, 家庭訪問評価尺度

摘要

家庭訪問評定量表最初是根據成功家訪的現場描述制定的, 並得到多學科的家訪訪問研究的支持。四個家庭訪問實踐量表包括與家庭建立關係的指標、對家庭優勢的響應、促進親子互動以及與家長的合作。三個家庭參與量表包括親子互動、父母參與和孩子參與訪問的指標。原始版本 HOVRS-1 使用來自兩個早期啟蒙方案家庭訪問計劃的視頻和數據進行了驗證。 HOVRS-3 的概念和結構變化旨在提高可讀性、可用性和清晰度。新訓練的觀察員使用 HOVRS-3 觀察原始測量樣本的存檔視頻。 HOVRS-3 表現出良好的評估者間可靠性、內部一致性、聚斂效度、預測效度、實際意義和版本穩定性。當 HOVRS-3 家庭訪問質量得分較高時, 育兒得分平均或更好, 兒童語言在3歲時達到適齡或更高水平, 超過1歲時的育兒和兒童語言。家庭訪問評定量表可以指導嬰兒及幼兒和早期兒童計劃中的家訪質量觀察, 以改善家庭訪問實踐和家庭參與。

關鍵詞

家庭參與, 家庭訪問實踐, 家訪質量, 家庭訪問評定量表

ملخص

تم تطوير مقاييس تقييم الزيارات المنزلية في البداية في ضوء الملاحظات الميدانية للزيارات المنزلية الناجحة والمدعومة من خلال أبحاث الزيارات المنزلية في تخصصات متعددة. وهناك أربعه مقاييس لممارسات الزيارات المنزلية وتشمل : مؤشرات بناء العلاقة مع الأسر ، والتعامل مع جوانب قوة الأسرة ، وتيسير التفاعل بين الوالدين والطفل ، والتعاون مع الوالدين. كما هناك ثلاثة مقاييس للمشاركة العائلية وهي: مؤشرات للتفاعل بين الوالدين والطفل ، ومشاركة الوالدين ، واندماج ومشاركه الأطفال في الزيارة. تم التحقق من صلاحية الإصدار الأول من مقياس تقييم الزيارة المنزلية (HOVRS-1) باستخدام ملفات فيديو وبيانات من برنامجين (هيد ستارت) للزيارة المنزلية. وتم إضافة تعديلات مفاهيمية وهيكلية على إصدار (HOVRS-3) لتحسين قابلية قراءته واستخدامه ووضوحه. وقد استخدم الملاحظون المدربون حديثا هذه النسخة (HOVRS-3) لمتابعة مقاطع الفيديو المؤرشفة من عينه القياس الأصلية. وأظهرت هذه النسخة موثوقية جيدة في القياس ، واتساق داخلي واسع ، وصلاحية متقاربة ، وصلاحية تنبؤية ، وأهمية عملية ، واستقرار في الإصدار. فعندما كانت درجات الجودة في الزيارات المنزلية أعلى طبقا لمقياس HOVRS-3 ، كان الاحتمال أكبر مرتين ان تكون درجات الابوة والأمومة متوسطه أو أفضل وأن تكون لغة الطفل في مستواه العمري أو أفضل في سن الثالثة. ويمكن لمقاييس تقييم الزيارات المنزلية أن تقدم التوجيه لملاحظات جوده الزيارة المنزلية في برامج الأطفال الرضع والطفولة المبكرة لتحسين ممارسات الزيارة المنزلية والمشاركة العائلية .

الكلمات الرئيسية

المشاركة العائلية ، الزيارات المنزلية ، جوده الزيارة المنزلية ، تقييم الزيارات المنزلية

1 | INTRODUCTION

A goal of most evidence-based home-visiting programs for families of infants and young children is to increase parenting behaviors that support children's early development (Sama-Miller et al., 2018). To have an impact on how parents support their children's development, home visitors are expected to engage parents in the program's intended process. Not surprisingly then, when Early Head Start (EHS) home visitors keep home visits more focused on child development, outcomes are significantly better for both parents and children (Raikes et al., 2006). Nevertheless, not all home visitors stay focused on child development or effectively engage parents in supporting child development. Variations in these aspects of home-visit practices and their effectiveness at engaging parents are related to corresponding differences in home-visiting outcomes for both parents and children (Roggman et al., 2016).

Observational research has identified specific home-visiting practices that are related to better parenting and child outcomes (Heinicke et al., 1999; Heinicke et al., 2000; Kelly, Zuckerman, & Rosenblatt, 2008; Peterson et al., 2018; Woods, Kashinath, & Goldstein, 2004; Zajicek-Farber, 2010): building positive trusting relationships, responding to families' strengths, encouraging positive parent–child interactions that support child development, and collaborating with parents as partners in the process. Family engagement in home visiting is similarly multifaceted, considering the engagement of the parent and child in home-visit activities and with each other during the home visit (Korfmacher et al., 2008; Wagner, Spiker, Inman Linn, Gerlach-Downie, & Hernandez, 2003).

Strategies such as observation, feedback, and encouragement appear to help home visitors increase positive, developmentally supportive parent–child interactions and collaborate with parents in a strengths-based manner to support the parent–child relationship (Fisher, Frenkel, Noll, Berry, & Yockelson, 2016; Moss et al., 2011; Van Doesum, Riksen-Walraven, Hosman, & Hoefnagels, 2008). Better parent and child outcomes are likely when home-visiting strategies attend to parents' strengths in parent–child interactions (Moss et al., 2011; Van Doesum et al., 2008; Zigler, Pfannenstiel, & Seitz, 2008), suggesting that measurement of the extent to which home visitors facilitate positive parent–child interaction is essential to better understand home-visiting processes that lead to intended outcomes for families.

Building positive, goal-oriented relationships with parents continues to be identified as an essential component of engaging families in home visiting and has impacts on the collaborations developed with parents to ensure that their needs are met (Axford, Lehtonen, Kaoukji, Tobin, & Berry, 2012). These collaborative relationships between home visitors and families can be viewed as a central mechanism for effective home visitors to have a positive impact on the families they serve. Effective home visitors value empathic, respectful, egalitarian relationships in which parent and home visitor can reflect on their experiences and learn together as life-long learners, using the family's strengths to move toward their goals for the child's and the family's well-being and development (Schaeffer, 2016). Measuring home visitor–family relationships is complex, however, and home visitors have noted that balancing program fidelity with meeting family needs can be challenging (Barak, Spielberger, & Gitlow, 2014). Furthermore, the home-visiting relationships are typically associated

with parents' engagement in home visits and home-visiting programs, and both have historically been difficult to define, distinguish, and measure (Korfmacher et al., 2008; Peterson et al., 2013).

2 | THE HOME VISIT RATING SCALES—ORIGINAL AND REVISED

2.1 | The original Home Visit Rating Scales

An objective measure of a home-visitor's practices and a family's engagement can help disentangle the complexity of the home-visiting process, estimate home-visit quality, and guide efforts to improve it. The Home Visit Rating Scales (HOVRS) comprise an observational measure of the quality of home-visit practices and family engagement. Research has supported the relationship-based, strengths-based, parent–child-oriented, equal partnership approaches to home visiting measured by the four practices scales of the HOVRS (cf. Roggman et al., 2016)—relationship building with families, responsiveness to family strengths and culture, facilitation of parent–child interaction, and collaboration with parents. The HOVRS also include three engagement scales—Parent–Child Interaction, Parent Engagement in the home-visiting process, and Child Engagement in the home-visiting process (Roggman et al., 2016), reflecting the importance of family engagement to effective home visiting (Raikes et al., 2006; Roggman, Boyce, Cook, & Jump, 2001; Wagner et al., 2003). Each of the seven scales is rated from 1 (needs support) to 7 (excellent) based on indicators at varying levels of quality. The scales were developed in partnership with EHS and other home-visiting programs in the course of several community research partnership projects and have been revised, adapted, expanded, reformatted, and reworded in various ways over time.

The original HOVRS (Roggman, Cook, Jump Norman, et al., 2008; Roggman et al., 2016), referred to here as the *HOVRS-1*, provided a set of quality ratings based on field-generated descriptions of successful home visits that also were supported by empirical research. The HOVRS-1 did not have explicit items but each scale instead listed varying numbers of indicators in four columns, increasing in quality from left to right: 1 (*inadequate*), 3 (*adequate*), 5 (*good*), and 7 (*excellent*). Observers checked whatever indicators they observed. Scale ratings, from 1 to 7, were made by the observer to reflect the pattern of indicators observed across the columns. The HOVRS-1 was validated with a sample of home-visit observations from 71 families in two EHS home-visiting programs, as part of a collaborative project examining home-visiting process from observations in a shared video archive (Roggman et al., 2016; Roggman, Cook, Jump Norman et al., 2008).

Our previous work showed that HOVRS-1 scales could be rated reliably by trained student researchers, was internally consistent within each scale, and demonstrated convergent and predictive validity (Roggman et al., 2016). In a sample of families from two EHS home-visiting programs, convergent validity was supported by significant correlations between HOVRS-1 total home-visit quality scores and independent measures reflecting home-visiting quality. HOVRS-1 scores were higher when home visits focused more on child development and families were more involved in the program. Predictive validity was supported by higher HOVRS-1 scores predicting better parenting environments and better child language development at the end of the home-visiting program. An indirect association of the HOVRS-1 total score with child outcomes via parenting outcomes was consistent with the aims of most home-visiting programs to improve child development by working through the parents to help them provide more everyday support for their children's development.

2.2 | Revisions of the HOVRS

The HOVRS-1 was subsequently adapted as the *HOVRS-A* (three quality levels) and the *HOVRS-A+* (four quality levels), in which the columns of indicators were organized into rows of parallel indicators within each scale, increasing in quality from left to right (Hallgren, Boller, & Paulsell, 2010; Roggman et al., 2012; Sparr, Korfmacher, Fulford, & Roggman, 2013). A series of further conceptual and structural changes increased clarity and consistency, resulting in the *Home Visit Rating Scales-3* (*HOVRS-3*; Roggman et al., unpublished) (see examples of these changes in Figure 1).

2.2.1 | Conceptual revisions

Conceptual changes clarify and differentiate the constructs that the HOVRS measures. Scale definitions in the HOVRS-3 describe the construct (e.g., "Facilitation of parent–child interaction: [Home visitor] Elicits and encourages positive, responsive, developmentally supportive caregiver–child interactions"). Item stems for each set of related indicators describe the specific aspect of the construct being observed (e.g., "Relationship Item 1: To show respect and acceptance of the family system, the home visitor …;" Parent "Engagement Item 6. To initiate activities and conversations, the caregiver …"). Wording is now more consistent and parallel between indicators within each item. Other wording changes more explicitly emphasize a relationship-based approach in the *Relationship* Scale, a strengths-based approach in the *Responsiveness* Scale, direct support of parent–child interaction in the *Facilitation* Scale, and collaborative process in the renamed *Collaboration* Scale (previously, *Nonintrusiveness)*.

The Relationship and Parent Engagement Scales are now more clearly differentiated by rewording several Relationship items to describe the home-visitor's relationship-building

Home Visitor Facilitation of Parent–Child Interaction							
Inadequate 1	2	Adequate 3	4	Good 5	6	Excellent 7	
Home visitor:							
__ Rarely helps parent respond to child's cues for interaction __ Ignores parent–child interactions __ Interacts with either parent or child but not both		__ Tries to facilitate interactions, even if not always effectively __ Tells parent to interact with child __ Tells child to		__ Facilitates some parent–child interactions __ Observes parent–child interactions and gives feedback __ Comments on child's cues for interaction		__ Consistently facilitates parent-child interactions __ Facilitates parent–child interactions that are rich and easy __ Provides appropriate suggestions and	

Overall: ☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Needs support		**Adequate**		**Good**		**Excellent**

1. To engage caregiver and child together, the home visitor
_ 1 = interacts with either the caregiver or the child but not both.
_ 3 = interacts with caregiver and child but occasionally directs attention to only one when possible to interact with both.
_ 5 = frequently interacts with both caregiver and child, excluding neither.
_ 7 = frequently interacts with both caregiver and child and helps sustain caregiver-child engagement.

2. To elicit ongoing caregiver-child interactions during this home visit, the home visitor
_ 1 = rarely tries to facilitate caregiver-child interactions.
_ 3 = occasionally tries to facilitate interactions, even if not always effectively.
_ 5 = consistently facilitates caregiver-child interactions.
_ 7 = consistently facilitates caregiver-child interaction and supports ongoing interaction, re-engaging as needed without interrupting.

FIGURE 1 Structural examples of the HOVRS-1 (top) and the HOVRS-3 (bottom) versions of the Home Visit Rating Scales

practices, instead of parent–home visitor interactions, and deleting two items (e.g., parent shares information, problems, or concerns openly with home visitor) that were highly correlated with items in the Parent Engagement Scale. These changes increased the utility of the HOVRS by focusing the first four scales on the home-visitors' practices. The family engagement scales reflect the effectiveness of home-visiting practices at engaging a particular family, but also are influenced by complex factors that limit the utility of family engagement alone as a direct measure of home-visit quality.

Other conceptual changes include rewording items in the Responsiveness Scale to emphasize a strengths-based approach of getting information about each family's strengths, interests, and goals for supporting the child's development and using that information together with parents to individualize home-visit content and process. The Responsiveness Scale now more clearly emphasizes adapting home-visiting services to each family by observing family strengths and planning topics and activities together with parents or other caregivers. The scales now use the term *caregiver* to refer to a mother, father, stepparent, grandparent, other kin, foster parent, or any other caregiver who may be the focus of home-visit services.

2.2.2 | Structural revisions

Structural changes increase the readability and usability of the HOVRS-3. Previous versions showed indicators in a grid structure that provided a visual display of quality, but required reading indicators in cells of wrapped text from one column to the next without an explicit description of the aspect of practice or engagement being observed. The HOVRS-3 is structured in a list format, with each scale listing four to seven descriptive items, and each item followed by four quality indicators—1 (*needs support*), 3 (*adequate*), 5 (*good*), and 7 (*excellent*)—formatted similarly to multiple-choice test items. In a list format, items and indicators are faster to read and compare than when in a row of cells. These changes have made it easier for new observers to learn the HOVRS-3. Finally, the HOVRS-3 scales were reordered to begin with the Relationship Scale. This change reflects the emphasis on relationship-based approaches in home visiting and the strengths of home visitors, who often score higher on this scale than on the other practice scales.

The HOVRS-3 retains the structure of seven scales: Relationship building with family (seven items, e.g., "To

discuss possibly sensitive issues respectfully and reflectively"); Responsiveness to family strengths and cultures (six items, e.g., "To adapt activities to the family's interests and needs"); Facilitation of caregiver–child interaction (six items, e.g., "To promote developmentally supportive interactions"); Collaboration with caregiver (five items, e.g., "To encourage the caregiver's ideas and interests for interactions with child"); Caregiver–child interaction (seven items, e.g., "To observe and respond to the child's behavior"); Caregiver engagement (six items, e.g., "To show interest in materials and activities"); Child engagement (four items, e.g., "To show interest and enthusiasm about home visit activities"). The structure of items and indicators allow scale ratings to be made by the observer or easily calculated from the item ratings. Similar to previous versions, a home-visit practices summary score can be averaged from the first four scales, an engagement summary score from the last three scales, and a total score from all seven scales.

2.3 | Purpose of study

This study examines the psychometric properties of the HOVRS-3 revision in relation to data from the original measurement sample. We reexamined the home-visit observations from the original measurement sample with the revised and restructured HOVRS-3. We used the new data from the HOVRS-3 to examine the psychometric properties of the revised version, including interrater reliability, scale reliability, convergent validity, and predictive validity, along with the practical significance in terms of what home-visit quality means for parent and child outcomes. We also examined version stability from the HOVRS-1 to the HOVRS-3.

3 | METHOD

To test the psychometric properties of the HOVRS-3, we used the original HOVRS-1 measurement development sample, training new student observers to make reliable HOVRS-3 ratings of home-visit quality for all available archived original VHS video recordings. We examined multiple psychometric properties of the HOVRS-3, including validity in relation to independently observed measures of home-visit quality, parenting outcomes, and child outcomes from the data archive of the Early Head Start Research and Evaluation Project (EHSREP; Administration for Children Youth, & Families, 2002). Video recordings of home visits were obtained as part of program fidelity evaluation methods and research protocols specific to each of two home-visiting program sites (see Roggman et al., 2016), but the videos were not originally intended for measurement development or to represent all EHS home visits.

3.1 | Observer training

Observers were six newly trained student researchers, undergraduate and graduate, who completed a brief online course about the HOVRS-3 and practiced scoring several videos of home-visit observations before testing their reliability. All observers met the criterion of 85% agreement on standard observations that had been rated by the measurement developers. Interrater agreement was maintained at 85% by having 23% of the observations ($n = 15$), randomly selected at regular intervals from the assigned English-language observations, independently rated by a second observer, without the first observer knowing which observations would be double-rated. Each observer was checked by a second observer proportionate to the number of observations that the first observer rated; for two observations, the second observer was the graduate student researcher who trained the observers and supervised the video scoring. Two observers were considered in agreement on an observation if their scale ratings were adjacent (within 1 point on a 7-point rating), their item ratings were adjacent (within 2 points on a rating of 1, 3, 5, or 7), and no more than three items within each scale differed (two items for the Child Engagement Scale, which has only four items). Before observing more home-visit videos, observers were required to review, discuss, and resolve any disagreements that did not meet the criteria.

3.2 | Measurement sample

The original HOVRS-1 measure was tested with recorded observations of home visits to 71 families (Roggman et al., 2016), a convenience sample based on available home-visit video and family-level data. All families had been enrolled at least 6 months before a home-visit observation and stayed enrolled at least 7 months, averaging 30 months of enrollment. Due to damaged VHS video recordings and other methodological barriers, videos from the original measurement sample could be rated with the HOVRS-3 for only 65 (92%) of the original families. Of the 65 families in our analysis sample, about one third (32%, 21 families) were from one site and the rest (68%, 44 families) were from a second site, both in semi-urban regions of the United States. For scale and rater-reliability estimates, all 65 cases were included. For validity testing, varying numbers of these families had data on other indicators of home-visit quality to examine convergent validity, and only 53 of them had outcome data available to examine predictive validity (80% of all families; 62% at Site 1, 89% at Site 2). Table 1 shows descriptive sample characteristics, noting any differences, for the measurement sample of 65 families and the validity subsample of 53 families with parenting and child outcome data.

Most families had only one home visit observed, but 6 families at Site 1 and 9 families at Site 2 had two visits observed,

TABLE 1 Characteristics[a] of the Home Visit Rating Scales measurement sample (N = 65) and validity subsample (n = 53)

Characteristics	Measurement sample n (%)	Validity subsample n (%)
Families		
2 visits coded and scores averaged	15 (23)	12 (23)
Enrolled prenatally	13 (20)	11 (21)
Received public assistance	11 (17)	9 (17)
3 or more risk factors[b]	22 (34)	21 (40)
Children		
Male	28 (44)	23 (43)
Firstborn	43 (66)	38 (72)
Qualified for disability services	8 (12)	8 (15)
Mothers		
At risk for depression	21 (32)	19 (36)
No high-school degree	17 (26)	15 (28)
Lived alone, without another adult[a]	15 (23)	14 (26)
English as a second language	8 (12)	4 (8)
White	54 (83)	48 (91)
19 or younger at child's birth	24 (37)	22 (42)
Unemployed and not in school	35 (54)	28 (53)

[a]No site differences except parent living alone without another adult (Site 1: 48%, Site 2: 12%); no significant differences between the validity subsample and other sample families except that all children who qualified for disability services were in the subsample.

[b]Family risk factors include mother under age 20 when the child was born, unemployed, living alone, not completing high school, and/or the family receiving public assistance.

so the ratings were averaged across the two visits. The number of home visitors in the home-visit observations differed by site (6 at Site 1 and 12 at Site 2), but information was not available in the data archive about individual home visitors. Generally, home visitors had caseloads of 10 to 12 families, used their personal vehicles to travel to each family's home every week for home visits that generally lasted well over 1 hr, and were trained to implement the *Head Start Program Performance Standards* for home-based programs (https://eclkc.ohs.acf.hhs.gov/archive/policy/prior-regulations/performance-standards-effective-until-102416). Home visitors ranged in age from the mid-20s to the mid-50s, were female, and had either a Child Development Associate credential, a bachelor's degree in early childhood or a related field, or extensive experience with infants and young children. Both sites had at least one bilingual Spanish-speaking home visitor.

3.3 | HOVRS-3 Measurement analysis

Reliability for the scale scores was examined for observers and scales. Interrater intraclass correlation coefficients (ICCs) were based on a subsample of 15 observations scored by at least two trained student observers. Although percent agreement between observer pairs was periodically checked to ensure accurate scoring before and during data collection, overall interrater reliability was estimated from 15 double-coded observations using ICCs. ICCs are more appropriate for interrater reliability than is the often-used κ statistic, which is more appropriate for categorical classifications and is distorted by high- and low-frequency scores (Feinstein & Cicchetti, 1990; Gwet, 2002). ICCs reflect the relative similarity of independent ratings across the full set of doubly scored observations and take into account the degree of disagreement (Hallgren, 2012). In addition, ICCs can be calculated based on the assumption that only one observer's ratings will be used as the measure for data analysis, the situation most likely for home-visit observations.

Scale reliability, or the cohesiveness among items within the same scale, was examined for all 65 observations using ICCs that assume all items are used for the scale score. The HOVRS-3 can be scored by reviewing the item ratings and using informed judgment guided by the pattern of item ratings, but some programs and researchers prefer to average the item ratings to derive the scale score. These calculated scale scores are highly correlated with the observer ratings, $r > .90$; and likewise, the summary scores based on calculated scores are correlated with those based on observer ratings, $r > .94$. We used observers' scale ratings for all the HOVRS-3 analyses, as we had for the HOVRS-1 (Roggman et al., 2016).

To explore other aspects of measurement validity, the HOVRS-3 scores were examined in relation to other independently measured indicators of home-visiting quality to test convergent validity and in relation to parent and child outcomes to test predictive validity. These indicators and outcomes were measured as part of the EHSREP data collection, for which the final report technical appendices fully describe the data-collection methodology and psychometric properties of the measures used (Administration for Children, Youth, & Families, 2002).

Convergent validity of the HOVRS-3 was examined in relation to available indicators of the quality of home visiting in EHSREP family-level measures of parent-reported enrollment, staff-rated program involvement, home-visitor-reported engagement, and home-visitor-reported content focus. Although the level of measurement and response format vary across these measures, they nevertheless provide indicators of home-visit quality used by other researchers. These four indicators of home-visiting participation and content are from different sources, at differing time points over varying durations, but they are indirect indicators of quality

that have been linked to stronger program outcomes in other studies (Raikes et al., 2006; Roggman, Cook, Peterson, & Raikes, 2008) and reflect home-visitors' effectiveness at engaging families, focusing on child development, and supporting family participation and retention.

Enrollment duration was measured by parent reports of home-visit dosage (how many periods a family received the intended number of home visits). *Family program involvement* was measured by staff ratings of family involvement in the overall program (Likert rating at exit). *Home visit engagement* was measured by home-visitor ratings of the primary caregiver's engagement in each home visit (reported per visit over approximately two years, averaged over visits). *Child development focus* of home visits was measured by home-visitor reports of the percent of home-visit time focused on child development versus family issues or relationship building (per visit over approximately two years, averaged over visits).

Predictive validity of the HOVRS-3 was examined using independent measures of parent and child outcomes, available from the EHSREP data archive, which showed key impacts of EHS in the national evaluation (Love et al., 2005). The *Home Observation Measure of the Environment* (*HOME*; Caldwell & Bradley, 1984) was used at child age 36 months, at the end of the EHS home-visiting program, and reflects the quality of developmental support for the child available in the parenting environment. The HOME is a robust measure that shows strong predictive validity for positive child outcomes across diverse cultures, ethnicities, and nationalities (Bradley, Corwyn, Burchinal, McAdoo, & Garcia-Coll, 2001). We examined the HOVRS-3 in relation to the HOME total score and the Language and Warmth subscale scores. The *Peabody Picture Vocabulary Test, Third Edition* (*PPVT-III*; Dunn & Dunn, 1997) was used to measure children's receptive vocabulary at 36 months, a critical foundation for later school success (McIntyre et al., 2017). The PPVT-III shows strong validity in relation to other established language-ability tests and was used to test the child-development impact of EHS in the national evaluation. We examined the HOVRS-3 in relation to the PPVT-III standard score ($M = 100$, $SD = 15$), based on national norms.

To examine how much difference quality makes for increasing children's chances of school success, the measure's practical significance, we examined the HOVRS-3 in relation to whether children were at age level on the PPVT-III, based on scores of 100 or greater, and in relation to whether their parenting environments were above average, as compared with other families in the same program, based on z-scores of 0 or greater. Both of these are stringent cutoffs, which define "above average" as a score at the mean or above. Earlier measures of parenting and child outcomes at 14 months were statistically controlled in these analyses by covarying the 14-month HOME (Caldwell & Bradley, 1984), a measure of the

TABLE 2 Mean and *SD* for Home Visiting Rating Scales ratings, summary scores, and total score (N = 65; 2 Early Head Start program sites)

Scores	M	SD
Home Visiting Practices Scales		
Scale 1. Relationship building with family rating	4.79	0.97
Scale 2. Responsiveness to family strengths rating	4.15	0.94
Scale 3. Facilitation of parent-child interaction rating	3.53	1.15
Scale 4. Collaboration with parent rating	3.96	1.13
Practices summary score (Scales 1–4)	3.91	0.88
Family Engagement scales		
Scale 5. Parent-child interaction rating	4.52	1.49
Scale 6. Parent engagement rating	4.56	1.20
Scale 7. Child engagement rating	5.22	1.36
Engagement summary score (Scales 5–7)	4.77	1.20

quality of the parenting environment during the infant–toddler period, and the 14-month *Communicative Development Index* (*CDI*; Fenson et al., 1994), a measure of children's early receptive vocabulary.

4 | RESULTS

4.1 | Description of scales

Mean ratings and *SD* of the mean for each HOVRS-3 scale are shown in Table 2. The pattern of high and low average ratings is similar to results from the HOVRS-1, with the Relationship Scale having the highest and the Facilitation Scale the lowest average ratings among the practices scales, and Child Engagement the highest among all the scales.

4.2 | Scale and rater reliability

Reliability of the HOVRS-3 was estimated with ICCs at the scale and observer levels (Table 3). At the scale level, ICCs reflect the interrelatedness of item ratings within the same scale and were calculated based on the assumption that all items are used for the overall scale rating. For all HOVRS-3 scales, ICCs were greater than .70 and for most scales greater than .80, reflecting good cohesiveness among the items within each of the HOVRS-3 scales. For the practices summary score (Scales 1–4), engagement summary score (Scales 5–7), and the HOVRS-3 total (Scales 1–7), ICCs were all above .75.

At the observer level, ICCs reflect the agreement between two observers and were calculated based on the assumption that the data from only one observer's ratings would be used as the measure, a higher standard than assuming that multiple

TABLE 3 Reliability estimates for Home Visit Rating Scales-3 (HOVRS-3): Scale ratings and summary scores

Scales and scores	No. of items	Interitem ICC[a] (N = 65)	Interrater ICC[b] (n = 15)
Scale 1. Relationship building with family	7	.81	.88
Scale 2. Responsiveness to family strengths	6	.71	.91
Scale 3. Facilitation of parent-child interaction	6	.82	.87
Scale 4. Collaboration with parent	5	.83	.74
Scale 5. Parent-Child Interaction	7	.94	.93
Scale 6. Parent Engagement	6	.86	.95
Scale 7. Child Engagement	4	.88	.97
Summary Score: Practices (Scales 1–4)	4	.76	
Summary Score: Engagement (Scales 5–7)	3	.86	
HOVRS-3 Total Score (Scales 1–7)	7	.87	

[a]This intraclass correlation (ICC; α) assumes all item ratings are averaged for use as data.

[b]This intraclass correlation assumes only a single-rater's scores are used as data.

observers' ratings would be averaged. Interrater ICCs, assuming a single observer, were all above .70 and for most scale ratings were greater than .85. These ICCs demonstrate good reliability between observers and show that ratings from only one observer are needed for reliable measurement data.

4.3 | Measurement validity

Measurement validity for the HOVRS-3 was tested in terms of convergent and predictive validity. Site differences were considered in all validity tests because the two sites differed significantly on the HOVRS-3 *practices score*, *engagement score*, and *total score*, along with independently measured *enrollment duration* and *child development focus*. All reported validity analyses control for site.

4.3.1 | Convergent validity

Convergent validity, showing that a measure of a construct is related to other measures of similar constructs, was tested by examining HOVRS-3 scores in relation to other home-visiting quality measures. The total and summary scores on the HOVRS-3 were tested in association with independent measures of home-visiting program quality: enrollment duration, program involvement, home visit engagement, and child

TABLE 4 Partial correlations, controlling for site, between Home Visit Rating Scales-3 (HOVRS-3) scores and other measures reflecting home-visiting program quality

Extant home-visit quality indicator	HOVRS-3 Practices score (Scales 1–4)	HOVRS-3 Engagement score (Scales 5–7)	HOVRS-3 Total score (Scales 1–7)
Enrollment duration	.15	.15	.17
Focus on child development	.10	.25*	.21
Engagement in program	.07	.27*	.20
Engagement in home visits	.20	.26*	.26*

*$p \le .05$.

development focus. The HOVRS-3 total score and engagement score were related to other measures of home-visiting program quality, supporting convergent validity (Table 4). The HOVRS-3 practices score, however, was not related to other measures of quality, approaching zero association with the family's overall involvement in all program activities, which could include participating in group activities for children and parents, social services case management, health and mental health services, program social events, and many other engagement opportunities unlikely to be influenced by home-visitor practices. In addition to the correlations on Table 4, the HOVRS-3 Responsiveness Scale was significantly correlated with enrollment duration, $r(48) = .32$, $p = .02$; the Parent Engagement Scale was significantly correlated with family program involvement, $r(51) = .35$, $p = .01$; and the Child Engagement Scale was significantly correlated with greater child development focus in home visits, $r(59) = .35$, $p = .005$.

4.3.2 | Predictive validity

Predictive validity, showing that a measure of a construct predicts what the construct is expected to predict, was tested by examining HOVRS-3 scores in relation to home-visiting outcome measures. The total and summary scores were examined in relation to program outcomes in the quality of the home environment that the parent provides for the child and the child's language ability.

The HOVRS-3 total score, practices score, and engagement score were significantly correlated with key parent and child outcome measures used in the evaluation of EHS (Administration for Children, Youth, & Families, 2002) when children were age 36 months, at the end of the home-visiting program: the parenting support for the child's development, as measured by the HOME (Caldwell & Bradley, 1984), and children's receptive vocabulary, as measured by the PPVT-III (Dunn & Dunn, 1997). Significant partial correlations, controlling for site, support the predictive validity of the HOVRS-

TABLE 5 Partial correlations, controlling for site, between Home Visit Rating Scales-3 (HOVRS-3) scores and program outcomes in parenting and child development

Outcome Variables	HOVRS-3 Practices score	HOVRS-3 Engagement score	HOVRS-3 Total score
HOME total–36 mo	.30*	.30*	.33*
HOME language–36 mo	.32*	.33*	.36*
HOME warmth–36 mo	.30*	.06	.19
HOME ≥ average–36 mo	.22†	.34*	.32*
PPVT-III–36 mo	.16	.26†	.23†
Child language ≥ age level–36 mo	.09	.32*	.24*

HOME = Home Observation Measure of the Environment; PPVT-III = Peabody Picture Vocabulary Test (3rd ed.); mo = months.
†$p \leq .10$.
*$p \leq .05$.

TABLE 6 Regression models testing the indirect association[a] of home-visiting total quality (Home Visit Rating Scales-3 [HOVRS-3]) with child language development (PPVT-III) at age 3 years, through parenting (HOME)

Predictors of HOME total environment	Model 1		
	B	SE	β
Site	−.32	.33	−.14
HOVRS-3 total	.36	.15	.34*
R^2 for model			.10*
F for R^2 change at last step			5.67*

Predictors of child vocabulary (PPVT-III)	Model 1			Model 2		
	B	SE	β	B	SE	β
Site	−.38	.35	−.16	−.24	.32	−.10
HOVRS-3 total	.27	.16	.26†	.12	.15	.12
HOME environment total				.43	.13	.44*
R^2 for model			.06†			.22*
F for R^2 change at last step			2.96†			10.74*

HOME = Home Observation Measure of the Environment; PPVT-III = Peabody Picture Vocabulary Test (3rd ed.).
[a]indirect effect test (HOVRS-3 to HOME to PPVT-III), Sobel test statistic = 1.93.*
†$p \leq .10$.
*$p \leq .05$.

3 in relation to these parent and child outcomes expected from home visiting (Table 5). Correlations with the HOME parenting scores appear more consistent than with the PPVT-III child language outcomes, as would be expected in a home-visiting model that takes an approach of working with parents to help them provide more support for their children's early development. Home-visiting practices were related primarily to parenting outcomes whereas family engagement was related to both parenting and child outcomes, although only the practices were significantly related to parent warmth. The HOVRS-3 total score predicted both parenting and child outcomes and was used for further analyses to explore indirect associations of home-visit quality in relation to parenting and child outcomes.

Early childhood home-visiting programs are thought to have impacts on child development through their effects on parenting and the developmental support that parents provide for their children (Raikes et al., 2014). To explore an indirect association of home-visiting quality with child language development through parenting support in the home environment, a series of multiple regression models, controlling for site, tested the indirect effect of the HOVRS-3 on child PPVT-III through the HOME (VanderWeele, 2015). As expected from the partial correlation results, the HOVRS-3 significantly predicted parenting scores on the HOME and approached significance on the PPVT-III child language scores, but the results also show a statistically significant indirect effect from the HOVRS-3 through the HOME to the PPVT-III, supporting a pathway from home visiting to parenting to child language. The practices and engagement scores showed similar indirect effects approaching significance, with each making a significant direct contribution to the HOME,

which then directly contributed to the PPVT-III. These indirect effects support a theory of change in which higher quality home visits, with both high-quality practices and strong family engagement, lead to better parenting environments, which in turn lead to better child outcomes (Table 6).

4.3.3 | Practical significance

Practical implications include whether the quality of home visits makes a meaningful difference in parenting and child-development outcomes. To consider a meaningful difference, we used logistic binary regression to test dichotomous outcomes reflecting average or better outcome scores, defined as at or better than the mean of the 36-month PPVT-III or HOME measures. We controlled for related earlier scores on the infant–toddler version of the HOME (Caldwell & Bradley, 1984) and the CDI measure of early receptive language development (Fenson et al., 1994), each measured at the 14-month data-collection point (Administration for Children, Youth, & Families, 2002). The HOVRS-3 significantly predicted better than average outcomes, regardless of site differences or Time 1 measures, on both the HOME and the PPVT-III (Tables 7 & 8). Odds ratios on Table 7 show that regardless of how good their home environments were at the beginning of the home-visiting program, families receiving higher quality home visits, across both sites, were over twice as likely to provide average or better parenting environments for their children,

TABLE 7 Logistic binary regression classifying above versus below average parenting environments from the Home Visit Rating Scales-3 (HOVRS-3), controlling for site and earlier parenting environment

Predictors of above average HOME–36 mo	B	SE	Wald	df	P-value	Odds ratio
Site	−.15	.87	.03	1	.86	.86
HOME–14 mo	.53	.22	5.70	1	.02	1.70
HOVRS-3 total	.81	.40	4.08	1	.04	2.25

HOME = Home Observation Measure of the Environment.

TABLE 8 Logistic binary regression classifying above versus below age-level child language from Home Visit Rating Scales-3 (HOVRS-3) score, controlling for site and earlier child language comprehension

Predictors of above average PPVT-III–36 mo	B	SE	Wald	df	P-value	Odds ratio
Site	1.91	1.21	2.49	1	.11	6.77
Child language–14 mo	.04	.03	2.92	1	.09	1.05
HOVRS-3 total	1.06	.49	4.58	1	.03	2.88

PPVT-III = Peabody Picture Vocabulary Test (3rd ed.); mo = months.

as compared with other families in the same program. Furthermore, odds ratios on Table 8 show that regardless of their early language abilities at the beginning of the program, children receiving higher quality home visits were almost three times more likely to have language skills at age level or better, scoring at the mean of 100 or higher compared with national norms, when they exited the program.

4.4 | Version stability

From the original to the current version, a correlation of .60 reflects adequate stability over multiple revisions. Figure 2

shows the average ratings in each scale from both the HOVRS-1 and the HOVRS-3 for the same set of observations. The average scale ratings from the HOVRS-3 are more similar across the scales and higher for several scales than are the ratings from the HOVRS-1.

Most of the conceptual revisions of the HOVRS-3 were in the Relationship Scales, which was refocused to emphasize home-visitor practices rather than considering the parent contribution to the parent–home visitor relationship, which is measured by the Parent Engagement Scales. The Relationship Scales ratings show more variability in the HOVRS-3 ratings than in the HOVRS-1, in which Relationship ratings had more limited range, smaller standard deviation, and higher kurtosis (2.53), with ratings of 3 or "good" for over two thirds of families' home visits. The Relationship Scales now reflects a wider range of quality in the relationship-building practices of the home visitor.

The Responsiveness Scales showed the greatest version change in mean score, with the HOVRS-3 averaging a full level of quality greater than the HOVRS-1, $t(64) = 9.94$, $p < .01$, increasing from "adequate" quality to about midway between "adequate" and "good" quality. The home visits are the same, but the overall low scores for this scale on the HOVRS-1 reflected specific, but rarely used, strategies rather than the more general strengths-based strategies described in the HOVRS-3 about how a home visitor responds to individual family strengths and interests by adapting home-visit content and process to better serve the family.

Overall, the scales show smaller SDs and more distributed ratings in the HOVRS-3 than in the HOVRS-1, but, with the exception of the Relationship Scale, were significantly correlated across versions and relatively stable over time. The engagement scales, which received the fewest conceptual changes in the revisions, changed very little in rating
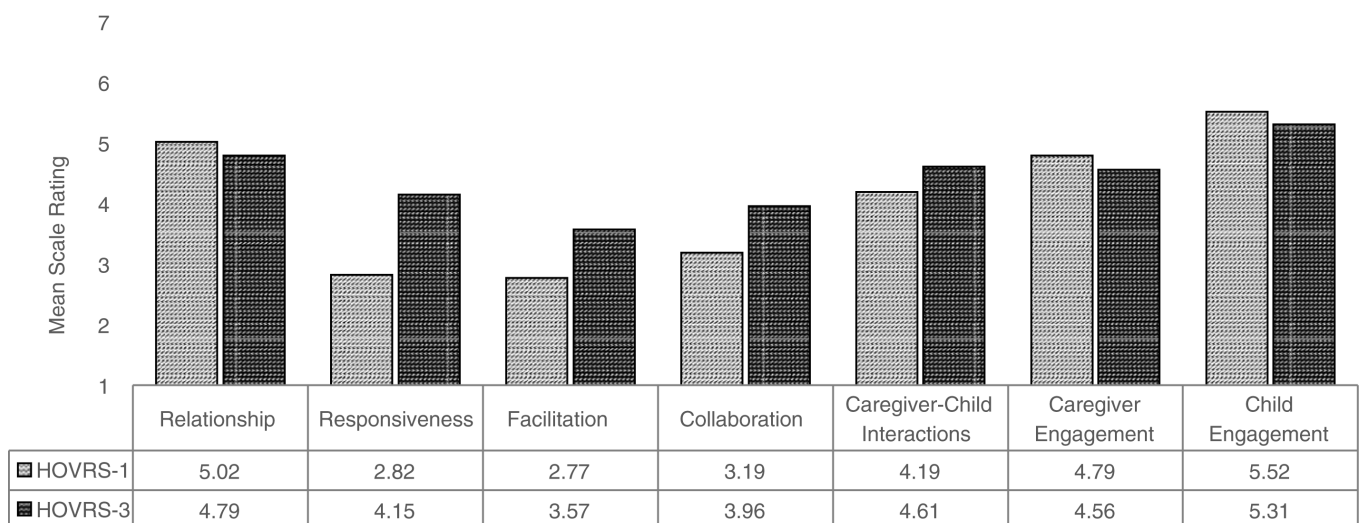

	Relationship	Responsiveness	Facilitation	Collaboration	Caregiver-Child Interactions	Caregiver Engagement	Child Engagement
HOVRS-1	5.02	2.82	2.77	3.19	4.19	4.79	5.52
HOVRS-3	4.79	4.15	3.57	3.96	4.61	4.56	5.31

FIGURE 2 Average scale ratings from the HOVRS-1 and the HOVRS-3 on the same measurement

variability from the HOVRS-1 to the HOVRS-3, and had the highest stability correlations.

5 | DISCUSSION

Our results support the reliability and validity of the HOVRS-3 as a measure of home-visit quality in home-visiting programs that focus on increasing parent support of children's early development. When home visits were rated as higher quality on the HOVRS-3, parent and child outcomes were better at the end of the program. High-quality home visits, as represented by higher scores on the HOVRS-3, were related to sharply greater chances for the parents to offer better than average home environments, as compared with other families who experienced the same program, and for the children to be at age level or better in early language development—a key to later school success. These results show that home-visiting quality is important for ensuring that home-visiting programs are effective at increasing positive outcomes for young children and their parents.

Consistent with the theory of change described by most EHS home-visiting programs (Raikes et al., 2014), variations in the quality of home visiting measured by the HOVRS-3 predicted child outcomes indirectly through the association of home-visiting quality with the developmental support parents provided in the home environment. These results align with evidence from the EHSREP study, showing that much of the impact of EHS home-visiting programs on children's development occurred indirectly through the impact on parenting (Administration for Children, Youth, & Families, 2002). The practices measured by the HOVRS, based on recommendations from multiple disciplines studying support services for families with infants and young children, were more strongly related to parenting than to child outcomes, as expected for home-visiting programs with theories of change based on working through the parent to the child. The total HOVRS-3 score was a more consistent predictor of both parenting and child outcomes than were either practices or engagement alone, particularly when controlling for earlier measures of parenting and child development, suggesting the use of the total score for assessing overall home-visiting quality. Together, these results strongly support a strategic pathway to child outcomes through parenting as a key mechanism of effective home-visiting interventions aiming to improve child-development outcomes.

The conceptual changes to the HOVRS-3, compared with the HOVRS-1, are evident primarily in the Responsiveness and Relationship Scales, which were both revised to reflect common home-visitor practices in adapting to and engaging families in the home-visiting process. Additional feedback from practitioners and supervisors in various home-visiting programs has indicated a need to further clarify descriptive indicators and use terms more consistently. These changes resulted in a wider range and more normally distributed scores on the HOVRS-3 scales and more similarity of average scores across the scales. These scales are now more congruent with research showing that home-visitor relationship skills are critical for home-visit effectiveness (Axford et al., 2012) and that adapting services to individual families contributes to family participation (Sweet & Appelbaum, 2004).

Structural revisions of the HOVRS-3 supported interrater reliability. We initially made the structural changes to guide students who were learning to collect data from home-visit observations with the HOVRS-3 and immediately received positive feedback from programs that had used earlier versions of the measure. We found that item stems helped observers understand the type of behavior that each set of indicators reflected. In addition, having each set of indicators structured in a multiple-choice list made them quicker to read and compare when rating each item. Our student observers for the HOVRS-3 were undergraduate or graduate students in psychology or child development who had an interest in services to families with young children and a commitment to learn research-based observational skills for rating the quality of home-visiting services in this and other sets of home-visit observations. These student researchers had some relevant education, but their observation skills were variable, and almost none of them had ever seen a home visit before they began observing videos in our observation laboratory. These characteristics made our observers similar to many new home visitors or program staff, and our observers' ability to readily learn to score the HOVRS-3 reliably supports the utility of the HOVRS by program staff who may be new to home visiting.

We believe the structural changes increase the utility of the HOVRS for home-visiting programs. Many kinds of home-visiting programs—EHS, Parents as Teachers, other evidence-based home-visiting program models, and newly emerging home-visiting models in the United States and other countries—have used the scales to observe home-visiting practices and family engagement. By having a readable, useful, and practical guide for observing home visits, supervisors, mentors, coaches, and consultants can use the HOVRS-3 to provide descriptive feedback to home visitors along with a clear indicator of the next step toward higher quality. Training and technical-assistance providers can use the HOVRS-3 to guide content for home-visiting practitioners. Program administrators can use the data to track program progress toward improving quality in their home visits. Home visitors can use the tool to guide self-reflection about the quality of a home visit. The HOVRS-3 will be useful for observation-guided reflection, supervision, mentoring or coaching, and continuous quality improvement in home visiting, along with articulating shared goals for home-visitors' professional preparation, in-service training, and professional development.

One strength and also a limitation of this study is testing the HOVRS-3 on the same home-visit observations as those of the HOVRS-1, thereby testing the psychometric properties of the revised measure on the same observations as the original measure. The advantage of doing this is that we were able to examine version stability, which is important for programs already using previous versions of the measure. Testing the HOVRS-3 on observations from the original measurement sample shows that a series of revisions has resulted in a version that remains highly correlated with the original version and maintains the reliability and validity of the original measure, but offers ratings with more conceptual clarity and statistical variability in a structure that is easier to learn and use. Additional research on other samples will be needed.

The small sample is another limitation of this study, as is the reliance on only two sites representing only a single evidence-based home-visiting model. The sample is limited not only in size but in diversity. Furthermore, the home visits were conducted almost 20 years ago in two relatively small EHS programs in two semi-urban internal regions of the United States. Fortunately, however, versions of the HOVRS-3 have been used recently in other home-visiting research projects, showing similar psychometric properties in larger, more diverse samples (Rubio Cordina, Dormal, & Araujo, this issue; Hughes-Belding et al., this issue; Manz & Ventresco, this issue). For further development of this and other measures of home-visiting quality, larger samples should be systematically selected from multiple home-visiting models across diverse communities and populations to allow more detailed analyses of items and indicators to increase rigor in the measurement of home-visiting quality. Additional research on home-visiting processes can identify new indicators of home-visiting quality, such as the triadic interactions identified by Hughes-Belding et al. (this issue), for testing convergent validity of this or other home-visit quality measures.

Monitoring, tracking, and improving the quality of home-visiting practices and family engagement has the potential to increase the effectiveness of home-visiting interventions used to reach multiple goals for families of infants and young children. Observing behavior, giving feedback, guiding reflection, and providing encouragement are not only essential skills for home visitors to use with parents (Moss et al., 2011; Van Doesum et al., 2008) but also are important skills for those who supervise and train home visitors. Home visits should be observed as often as classrooms to assess the quality of practices and the engagement of families and to guide coaching to improve home-visiting practices. Observational measures of the quality of practices in classrooms serving infants and young children are recommended for coaching teachers, monitoring classroom quality, and identifying teacher training or support needs (Isner et al., 2011). The revisions to the HOVRS have increased its utility for providing detailed feedback on the quality of home-visiting practices and engagement that can be used for guiding the professional development of home-visiting practitioners and informing the ongoing quality-improvement efforts of home-visiting programs.

Whether used to deliver early childhood services, early intervention, or maternal–child health services, home-visiting services for families with children 0 to 5 years of age often occur at a distance from central office support staff, making support and feedback challenging to provide even though critical for home-visitor effectiveness. The HOVRS can guide regular observations of home-visit quality to ensure that home visitors and families receive the support they need to be able to successfully collaborate in ways that improve outcomes for children.

CONFLICT OF INTEREST

The authors of this paper are also authors of the published Home Visit Rating Scales and provide training and consulting on the measure.

ORCID

Lori A. Roggman https://orcid.org/0000-0002-0048-9518
Gina A. Cook https://orcid.org/0000-0001-9763-6002
Mark S. Innocenti https://orcid.org/0000-0002-4911-5170
Vonda K. Jump Norman
https://orcid.org/0000-0002-6184-4819
Lisa K. Boyce https://orcid.org/0000-0003-2837-1332
Tasha L. Olson https://orcid.org/0000-0002-2296-0198
Katie Christiansen
https://orcid.org/0000-0001-6710-9172
Carla A. Peterson https://orcid.org/0000-0002-6094-0198

REFERENCES

Administration for Children, Youth, & Families. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start*. Washington, DC: Department of Health and Human Services.

Axford, N., Lehtonen, M., Kaoukji, D., Tobin, K., & Berry, V. (2012). Engaging parents in parenting programs: Lessons from research and practice. *Children and Youth Services Review*, *34*(10), 2061–2071. https://doi.org/10.1016/j.childyouth.2012.06.011

Barak, A., Spielberger, J., & Gitlow, E. (2014). The challenge of relationships and fidelity: Home visitors' perspectives. *Children and Youth Services Review*, *42*, 50–58. https://doi.org/10.1016/j.childyouth.2014.03.023

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H. P., & Garcia-Coll, C. (2001). The home environments of children in the United States Part II: Relations with behavioral development from birth through age 13. *Child Development*, *72*, 1868–1886. https://doi.org/10.1111/1467-8624.t01-1-00383

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock.

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-l

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185. https://doi.org/10.2307/1166093

Fisher, P. A., Frenkel, T. I., Noll, L. K., Berry, M., & Yockelson, M. (2016). Promoting healthy child development via a two-generation translational neuroscience framework: The Filming Interactions to Nurture Development video coaching program. *Child Development Perspectives*, *10*(4), 251–256. https://doi.org/10.1111/cdep.12195

Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment*, *1*(6), 1–6.

Hallgren, K., Boller, K., & Paulsell, D. (2010). *Partnering with families for early learning home visit observations: Better beginnings*. Princeton, NJ: Mathematica Policy Research.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Heinicke, C. M., Fineman, N., Rodning, C., Ruth, G., Recchia, S., & Guthrie, D. (1999). Relationship-based intervention with at-risk mothers: Outcome in the first year of life. *Infant Mental Health Journal*, *20*, 349–374. https://doi.org/10.1002/(SICI)1097-0355(199924)20:4<349::AID-IMHJ1>3.0.CO;2-X

Heinicke, C. M., Goorsky, M., Moscov, S., Dudley, K., Gordon, J., Schneider, C., & Guthrie, D. (2000). Relationship-based intervention with at-risk mothers: Factors affecting variations in outcome. *Infant Mental Health Journal*, *21*(3), 133–155. https://doi.org/10.1002/1097-0355(200007)21:3<133::AID-IMHJ1>3.0.CO;2-P

Hughes-Belding, K., Peterson, C. A., Walter, M. C., Rowe, N., Fan, L., Dooley, L. J., … Goodman, K. (2019). Quality home visits: Activities to promote meaningful interactions. *Infant Mental Health Journal*, *40*(3), 331–342.

Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L., & Burkhauser, M. (2011). *Coaching in early care and education programs and quality rating and improvement systems (QRIS): Identifying promising features*. Bethesda, MD: Child Trends.

Kelly, J. F., Zuckerman, T., & Rosenblatt, S. (2008). Promoting first relationships: A relationship-focused early intervention approach. *Infants & Young Children*, *21*, 285–295. https://doi.org/10.1097/01.iyc.0000336541.37379.0e

Korfmacher, J., Green, B. L., Staerkel, F., Peterson, C., Cook, G., Roggman, L. … Schiffman, R. (2008). Parent involvement in early childhood home visiting. *Child & Youth Care Forum*, *37*, 171–196. https://doi.org/10.1007/s10566-008-9057-3

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., … Fuligni, A. S. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, *41*(6), 885–901. https://doi.org/10.1037/0012-1649.41.6.88

Manz, P. H., & Ventresco, N. E. (2019). Observing home-visiting quality across time: A longitudinal reliability study of the Home Visit Rating Scales. *Infant Mental Health Journal*, *40*(3), 363–379.

McIntyre, L. L., Pelham, W. E., Kim, M. H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2017). A brief measure of language skills at 3 years of age and special education use in middle childhood. *Journal of Pediatrics*, *181*(2), 189–194. https://doi.org/10.1016/j.jpeds.2016.10.035

Moss, E., Dubois-Comtois, K., Cyr, C., Tarabulsy, G. M., St-Laurent, D., & Bernier, A. (2011). Efficacy of a home-visiting intervention aimed at improving maternal sensitivity, child attachment, and behavioral outcomes for maltreated children: A randomized control trial. *Development and Psychopathology*, *23*(1), 195–210. https://doi.org/10.1017/s0954579410000738

Peterson, C. A., Hughes-Belding, K., Rowe, N., Fan, L., Walter, M., Dooley, L., … Steffensmeier, C. (2018). Triadic interactions in MIECHV: Relations to home visit quality. *Maternal and Child Health Journal*, *22*(2), 1–10.

Peterson, C. A., Roggman, L. A., Green, B., Chazan-Cohen, R., Korfmacher, J., McKelvey, L., … Atwater, J. (2013). Home visiting processes: Relations with family characteristics and outcomes. *ZERO TO THREE Journal*, *33*(3), 39–44. https://doi.org/10.1007/s10995-018-2534-x

Raikes, H., Green, B. L., Atwater, J., Kisker, E., Constantine, J., & Chazan-Cohen, R. (2006). Involvement in Early Head Start home visiting services: Demographic predictors and relations to child and parent outcomes. *Early Childhood Research Quarterly*, *21*, 2–24. https://doi.org/10.1016/j.ecresq.2006.01.006

Raikes, H. H., Roggman, L. A., Peterson, C. A., Constantine, J., Brooks-Gunn, J., Chazan-Cohen, R., … Shiffman, R. (2014). Theories of change and outcomes in Early Head Start home-based programs. *Early Childhood Research Quarterly*, *29*(4), 574–585. https://doi.org/10.1016/j.ecresq.2014.05.003

Roggman, L. A., Boyce, L. K., Cook, G. A., & Jump, V. K. (2001). Inside home visits: A collaborative look at process and quality. *Early Childhood Research Quarterly*, *16*, 53–71. https://doi.org/10.1016/S0885-2006(01)00085-0

Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V., Boyce, L. K., Christiansen, K., & Peterson, C. A. (2016). Home visit quality variations in two Early Head Start programs in relation to parenting

and child vocabulary outcomes. *Infant Mental Health Journal*, *37*(3), 193–207. https://doi.org/10.1002/imhj.21565

Roggman, L. A., Cook, G. A., Jump Norman, V. K., Christiansen, K., Boyce, L. K., & Innocenti, M. S. (2008). The Home Visit Rating Scales (HOVRS). In L. A. Roggman, L. K. Boyce, & M. S. Innocenti, *Developmental parenting: A guide for early childhood practitioners* (pp. 209–217). Baltimore, MD: Brookes.

Roggman, L. A., Cook, G. A., Jump Norman, V. K., Innocenti, M. S., Christiansen, K., & Boyce, L. K., with Aikens, N., Boller, K., Paulsell, D., & Hallgren, K. (2012). *Home Visit Rating Scales-Adapted & Extended (HOVRS-A+)*. Unpublished measure.

Roggman, L. A., Cook, G. A., Peterson, C. A., & Raikes, H. H. (2008). Who drops out of Early Head Start home visiting programs? *Early Education and Development*, *19*(1), 1–26. https://doi.org/10.1080/10409280701681870

Rubio-Cordina, M., Dormal, M., & Araujo, M. C. (2019). Observing home-visiting quality at scale with the Home Visiting Rating Scales and a supervisor checklist in Peru. *Infant Mental Health Journal*, *40*(3), 343–362.

Sama-Miller, E., Akers, L., Mraz-Esposito, A., Zukiewicz, M., Avellar, S., Paulsell, D., & Del Grosso, P. (2018). *Home visiting evidence of effectiveness review: Executive summary*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Schaefer, J. K. (2016). Personal characteristics of effective home visitors. *Journal of Social Service Research*, *42*(1), 84–95. https://doi.org/10.1080/01488376.2015.1078868

Sparr, M., Korfmacher, J., Fulford, J., & Roggman, L. (April, 2013). *Assessing home visiting program quality at the point of service*. Paper presented at the meeting of the Society for Research in Child Development, Seattle, WA.

Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, *75*(5), 1435–1456. https://doi.org/10.1111/j.1467-8624.2004.00750.x

VanderWeele, T.J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York, NY: Oxford University Press.

Van Doesum, K. T., Riksen-Walraven, J. M., Hosman, C. M., & Hoefnagels, C. (2008). A randomized controlled trial of a home-visiting intervention aimed at preventing relationship problems in depressed mothers and their infants. *Child Development*, *79*(3), 547–561. https://doi.org/10.1111/j.1467-8624.2008.01142.x

Wagner, M., Spiker, D., Inman Linn, M., Gerlach-Downie, S., & Hernandez, F. (2003). Dimensions of parental engagement in home visiting programs: Exploratory study. *Topics in Early Childhood Special Education*, *23*(4), 171–187. https://doi.org/10.1177/02711214030230040101

Woods, J., Kashinath, S., & Goldstein, H. (2004). Effects of embedding caregiver implemented teaching strategies in daily routines on children's communication outcomes. *Journal of Early Intervention*, *26*, 175–193. https://doi.org/10.1177/105381510402600302

Zajicek-Farber, M. L. (2010). Building practice evidence for parent mentoring home visiting in early childhood. *Research on Social Work Practice*, *20*(1), 46–64. https://doi.org/10.1177/1049731509333172

Zigler, E., Pfannenstiel, J. C., & Seitz, V. (2008). The Parents as Teachers program and school success: A replication and extension. *Journal of Primary Prevention*, *29*(2), 103–120. https://doi.org/10.1007/s10935-008-0132-1