Summary of Gaps by Research Quality Questions and Methodology Factors.

Quality Indicator		Weaknesses/Areas of Improvement
Evidence that the Research Included and was Implemented by Low SES Participants in Authentic Settings		
I.	Participating organizations and settings	1. Only 26% of all identified language intervention studies generated evidence likely to generalize to other low SES participants. 2. Community settings such as libraries, grocery stores, for example, were rarely included in extant intervention research.
II.	Participants' characteristics	3. Compared to the frequently reported child sociodemographic characteristics, adult characteristics were often incomplete or not reported.
III.	Settings	4. Only 17 of the 140 studies (12 %) of studies investigated the effects of combining interventions occurring in more than one setting within the child's language environment.
IV.	Intervention implementers	5. The majority of studies reported that research assistants (staff) rather than parents and/or child care providers delivered the intervention to children, particularly so in the nonparent group. 6. Some adults who would be expected implementers in a scaled up Word Gap prevention effort (i.e. fathers, child care providers, and adoptive parents) have had limited or no intervention research involvement.
Evidence that the Intervention Happened as Expected		
V.	Fidelity of implementation measurement	7. Less than half of studies reported assessing fidelity of implementation. The most complete fidelity data reports were in SCD design studies. 8. When fidelity outcomes were reported, a majority of studies indicated that fidelity was less than adequate or unclear.
VI.	Language environment measurement	9. Less than half of studies observationally assessed the child's language environment, however, it was most often assessed in the parent implementer group.
Evidence that the Intervention was Working		
VII.	Research design	10. Twenty-seven percent of the literature is based on findings from QED designs that are subject to selection bias due to non-equivalent groups. While matching of participants was the strongest control for selection bias in QED designs, one's confidence is always suspect because matching procedures may be imperfect or incomplete. 11. The frequent use of QEDs and SCDs in the literature (42%) rather that the more rigorous RCT points to the greater challenges faced in conducting RCTs. QEDs and SCDs are often easier and cheaper to employ and participants are not nearly as reluctant because they do not face the uncertain outcome of randomization to treatment.

		12. The frequent use of QEDs and SCDs in the literature may also suggest that the questions being asked are pilot in nature, regarding development and refinement of new treatments. 13. The counterfactual conditions used in more than half of the studies were "business as usual", rather than alternative treatments, suggesting that the state of our work appears to be on developing evidence-based practices, rather than comparing the impacts of several evidence-based alternative treatments. 14. Attrition of up to 25% or more was a threat in the RCT and QED studies. Attrition was not a large issue in SCD studies. 15. The majority of intervention durations in all three design was only 2-8 weeks long. 16. The majority of follow-up research questions were focused on short- versus long-term effects.
VIII.	Child outcome measurement	17. Authors routinely failed to report whether or not findings were subject to measurement bias (i.e., assessor's blindness to treatment groups). While the RCT group was better at reporting this information, the QED and SCD groups were least forthcoming. 18. Less than half of studies reported reliability information for the measures they used, fewer reported validity information.
IX.	Statistical analysis	19. Only half of all the studies reported effect sizes. Effects sizes were most frequently reported for changes in children's outcomes, followed by parent and EC personnel outcomes. 20. Effect sizes were reported more frequently in the RTC studies, compared to the QED, and SCD studies 21. On average, nonparent implementation group studies reported larger effect compared to the parent and the P+NP groups, on the order of a .10 of a standard deviation difference. 22. Only five SCD studies reported using effects sizes and only one used an overlap effect size for single case data (Tau-U). 23. A weakness in the treatment for missing data was removal of missing cases rather than imputing or estimating missing data in analyses (Enders, 2010).
X.	Study limitations and recommendations reported	24. A majority of studies reported limitations concerned with study invalidity (internal and external validity) issues, and addressing issues for improving future research. 25. Even in this research with low SES participants, issues related to the suitability and effectiveness of the intervention with diverse children and families was an issue. 26. Studies reported implications primarily focused on future research, only moderately addressed practice implications, and rarely addressed policy implications.
Evidence about Why or How the Intervention was Working		
XI.	Moderator/ mediator	27. Only a third of studies reported addressing questions and conducting analyses of the influences of variables other than

	analyses	the intervention of study outcomes (moderator or mediator analyses). 28. Measuring only child outcome was a limitation in moderator/ mediator analyses, where additional information is needed, for example, the child's language environment, caregiver-child interactions, treatment fidelity, and adult outcomes.
XII.	Adult outcome measurement	29. Unlike child outcomes, parent/family outcomes measures were not universally assessed and reported in P group studies, and staff outcomes in NP studies.
Evidence that the Intervention was Accessible and Usable at Scale		
XIII.	Teaching adults to use the intervention strategies	30. Twenty-one percent of studies lacked information on the methods used to deliver the intervention to adults designed to change their communication with the child, particularly in the nonparent implementers group. 31. Another weakness was the frequent reliance on only one delivery method, for example, group training when an extensive professional development literature indicates that group training alone without ongoing implementation supports is not effective. 32. Several promising adult teaching procedures have yet to receive sufficient focus and development, for example, the use of adult peers as coaches or trainers of other adults. 33. The effects of linking information campaigns to promoting adult behavior change has not yet been examined.
XIV.	Social validity	34. Participants' reports of the social importance and acceptability of treatment goals, acceptability of procedures, and outcomes (social validity) was assessed infrequently overall but more likely in SCD studies. 35. When assessed, adult intervention implementers did not universally rate their experiences highly.
XV.	Digital technology	36. Results indicated that use of digital technologies is only beginning to support scalability, its largest role in this literature was in data collection.

Table Reference

Enders, C. K. (2010). *Applied Missing Data Analysis*. New York: The Guilford Press.