User-Informed Principles

Developing Assessments for All Early Learners

MEASURES FOR EARLY SUCCESS

Supporting Early Learners and Educators with Innovative, Equitable Assessments

Introduction 3

User-Informed Principles

Goal 1: Content 9

Goal 2: Psychometrics 15

Goal 3: Experience 19

Goal 4: Usefulness 23

Goal 5: Scalability 29

Appendices 34

Introduction

The Measures for Early Success Initiative, with funding support from the Bill & Melinda Gates Foundation, aims to reimagine the landscape of early learning assessments for the millions of 3- to 5-year-olds enrolled in Pre-K,[1] so that more equitable data can be applied to meaningfully support and strengthen early learning experiences for all young children.[2]

The present document outlines design parameters for child assessments that meet these outstanding needs in the field of early childhood education. This Target Product Profile (defined in more detail below) is intended to spur critical investments and innovations in the existing Pre-K assessment landscape, and support important stakeholder collaborations among Pre-K decision-makers, assessment suppliers, researchers, educators,[3] and families with children in Pre-K to develop innovative assessment solutions. This document is designed to be sufficiently ambitious to meet the needs in the field for more equitable, scalable, and useful measures of children's skills and competencies.

[1] Head Start Early Childhood Learning & Knowledge Center (2019); The National Institute for Early Education Research (2022); Home Grown (2020).

[2] Pre-K settings vary across states and in mixed-delivery systems; they may include public schools, child care, Head Start, and home-based child care.

[3] Educators refer to adults who are responsible for the care and education of children in Pre-K settings.

Our Strategy

Research shows that high-quality Pre-K can have lasting impacts on children's learning and development, setting them on a path for success in kindergarten and beyond, with the largest effects being clustered among children of color, emergent bilinguals, and children experiencing poverty. Frequent, systematic measurement of children's skills and competencies can help educators and families understand children's strengths and identify areas for growth, so that they can be responsive to their needs. Likewise, such information can help policymakers and system administrators make data-informed decisions about how to improve the quality of future Pre-K settings and programming. Yet policymakers, administrators, and educators often wrestle with a dearth of accurate, reliable, and timely data about children's skills, knowledge, and competencies in Pre-K settings that is routinely and systematically captured on a large scale. This lack of information presents an opportunity to reimagine how assessments can provide educators with the necessary information to enhance the quality of Pre-K settings across the United States.[4]

The Measures for Early Success Initiative aims to collaborate with critical stakeholders such as educators, families, and system leaders to create the next generation of assessments. By enhancing the availability of assessments of children's early learning that are useful, are scalable, and generate high quality information, the initiative's goal is to provide educators, families, and systems with accurate, reliable, and timely information to support the growth and development of all children in reaching their full potential.

To this end, these assessments must be:

— inclusive, equitable, and culturally and linguistically responsive;

— joyful, play-based, and developmentally appropriate for Pre-K aged children;

— embedded into typical Pre-K activities;

— easy to implement in diverse Pre-K settings; and

— supportive of teacher-child interactions.

The assessments must also:

— be adaptive to meet children where they are in their learning, development, and competencies;

— generate accurate, reliable, and timely information on children's growth and development multiple times per year;

— be psychometrically valid and predictive of positive outcomes for young children from diverse backgrounds;

— be intuitive for educators to train on and integrate easily into the Pre-K setting;

— create actionable and timely information for educators;

— be accessible and useful for families;

— leverage technology in a developmentally appropriate way; and

— be affordable and effectively used by Pre-K programs that serve communities with fewer resources.[5]

[4] Hsueh (2021).

[5] Affordability refers to the costs of purchasing assessments, training and supporting staff to administer them, meeting technological needs associated with them, and interpreting the data resulting from them.

Our Strategy *(continued)*

With this universal goal in mind, this Target Product Profile prioritizes the experiences, strengths, and needs of Pre-K educators, families, and children whose perspectives are less often elevated in the early design, creation, and validation of measurement and assessment tools—in particular, Black and Latine children,[6] children who speak Spanish and English, and children experiencing poverty. These groups are incredibly heterogeneous—and have unique experiences in the United States.[7] While these groups do not reflect the experiences of *all* historically marginalized groups of children, they are an important starting point for ensuring the development of more equitable and inclusive tools that meet the needs of a broad range of families and children being served in publicly funded Pre-K programs today. As such, the design parameters outlined in this document reflect what research and best practices suggest are needed, and highlights areas for further innovation, pilot-testing, and iteration on the assessments in collaboration with a broad range of stakeholders. In the future, the initiative will aim to draw upon these insights to consider and expand the design of assessment tools toward other groups of children.

---

[6] Consistent with experts in the field, this document uses the gender-neutral term "Latine" to refer to individuals whose cultural background originated in Latin America. In U.S. academic circles, Latinx is often used as a gender-inclusive term to refer to people from Latin American backgrounds, but Spanish-speakers find that Latinx is unpronounceable in Spanish. Therefore, we have opted to use the term Latine, which is commonly used throughout Spanish-speaking Latin American countries (see Melzi, McWayne & Ochoa, 2022).

[7] These prioritized populations are not a monolith and we expect that considerable heterogeneity will exist even within these groups that result from the many ways in which these categories might intersect (e.g., across race and income, or those who are monolingual speakers vs. those who are learning English and Spanish). These intersections will have important implications for how assessment items and solutions are constructed. In addition, poverty should be addressed as a structural barrier that can also manifest differently across ecologies (e.g., immigration status, intergenerational poverty, rural and urban poverty). Future assessments will need to account for both variations and commonalities in how learning and developmental constructs appear across children's varied social and cultural contexts in Pre-K settings.

What Is a Target Product Profile?

A Target Product Profile is a document that has been historically written for the global health community to articulate goals, requirements, and specifications for development of health care solutions. The Measures for Early Success Initiative has adapted this approach to support the development of assessment solutions for Pre-K. This document is composed of user-informed principles that identify key goals of a proposed child assessment solution with corresponding criteria and target thresholds that developers of assessment tools will work to attain.

How Can This Document Be Used?

This document will help product development teams that wish to build new assessment tools that can address the challenges of today's assessment landscape in the following ways:

- **Development of measurement items or solutions.** To identify measurement items or solutions that would assist in achieving outlined goals.
- **Gap analysis.** To identify places in which existing solutions do not address the goals outlined.
- **Prioritization.** To identify priority areas/features in the context of a product development roadmap.
- **Equitable design.** To develop features and address gaps and/or priorities with culturally responsive solutions and approaches.
- **Solution ideation.** To inform ideation sessions in which product teams generate assessment solutions via a number of possible methods including co-design sessions, workshops, and team exercises.

For institutional decision-makers that seek to adopt new child assessment tools, this document can serve to highlight important principles for evaluating and purchasing new solutions on the market that meet or exceed the thresholds for performance outlined in this document.

How This Document Should *Not* Be Used

- The user-informed principles provided in this document should not be viewed as static or overly prescriptive. Indeed, this document does not articulate the process by which new child assessments should be created. Instead, this document provides the structure by which a more open, creative, and iterative approach can be set in motion, one that seeks to address gaps in research, and promote more equitable processes for ideation and testing in the field.
- New assessment products produced for the market should not be viewed as a solution for long-standing, educational inequities but rather one of many important levers used to improve learning and developmental outcomes for all young children.
- The user-informed principles in this document should not be used to create additional high-stakes tests for young children; rather the user-informed principles aim to support the development of assessment tools that will help educators and programs with the ongoing improvement of Pre-K experiences for all young children.

Development Methodology

The user-informed principles provided in this document were developed by MDRC and Substantial, with funding support from the Bill & Melinda Gates Foundation's Early Learning Team and with the contributions and engagement of educators, families, program administrators, state and local Pre-K system administrators, academic experts, and innovators and entrepreneurs in the Pre-K field. Data sources included interviews, focus groups, ideation workshops, and a literature review. Across these groups, we prioritized engagement of contributors in Head Start, community-based child care, and other publicly funded Pre-K settings that support children of color, emergent bilinguals, and children who experience poverty. More specifically, the groups engaged included:

- Black and Latine families, as well as families currently experiencing poverty, with a child who had recently completed a year of Pre-K in primarily publicly funded programs;
- Pre-K educators, instructional coaches, and administrators who primarily serve the prioritized populations;
- Academic and Pre-K operations experts, consisting of assessment researchers, developmental psychologists, nonprofit leaders, heads of agencies serving young children, experts on emergent bilinguals (specifically learners of Spanish and English), and family socialization and learning experts across Black and Latine communities; and
- An array of Pre-K decision-makers, advocates, and thought leaders from more than 20 geographically diverse states, localities, and Pre-K program operators.

Overview of User-Informed Principles

The following are thematic areas to address in future-state assessments:

1. **Content**—pertains to children's development, skills, and competencies measured within domains of early learning.
2. **Psychometrics**—the extent to which assessment(s) reflect valid, psychometrically sound, and comparable results for children across races, ethnicities, income levels, early childcare settings, or geographic areas within the United States.
3. **Experience**—reflects the optimal experience of important stakeholders using or benefiting from the assessment(s) and the resulting data, including children, educators, and families.
4. **Usefulness**—the meaningful utilization of the assessment data and findings for important stakeholders, including educators, families, Pre-K programs, and policymakers. Data will inform how educators can tailor children's support and instruction appropriately, and also inform program and policy decisions.
5. **Scalability**—the degree to which assessment(s) can be readily expanded in their use, and updated with research developments and/or adapted to local settings.

Description of Tables

In the following sections, the principles are specified in tables outlining key pieces of information:

1. Principles with intended goal.
2. More detailed subgoal for each goal.
3. Criteria that list the general design parameters for the subgoal identified.
4. An aspirational target threshold that provides indicators or suggested practices to assess the degree to which subgoals are successfully met.

The goals specified for each principle are designed to be aspirational. Thus, there may be tensions that arise in trying to achieve all the features outlined below. However, unique opportunities for innovation and breakthrough solutions may be possible if product development teams seek to address multiple goals holistically.

GOAL

1

Content

Instrument(s) comprehensively measure the skills and development of 3– to 5–year–old children in equitable and culturally responsive ways.

All children, irrespective of demographic background, are engaged in rich traditions of language and literacy as well as exploring, learning, and organizing their worlds, long before they start their formal education. Young children's knowledge about the world around them is shaped by the richness of their experiences and interactions with others and the cultural and social contexts in which they develop.[8] While there have been many research studies that have documented the unique skills and competencies of young children from racially, ethnically, linguistically, and socioeconomically diverse backgrounds,[9] many of them have not been measured on a wide scale or integrated into commonly used child assessments.[10] These skills and competencies may reflect universal constructs that are culturally and linguistically relevant across a broad array of communities of children. Further, they may be developmentally relevant, malleable constructs that are meaningful for setting young children on a path for success in kindergarten and beyond. Researchers as well as families can be actively engaged to extend what we know about children's socialization, learning, and development—particularly within the prioritized populations—and contribute in an ongoing manner to the assessment development process.

---

[8] Garcia Coll et al. (1996).

[9] For an example, see: Gardner–Neblett & Iruka (2015).

[10] Garcia Coll et al. (1996); Peña & Halle (2011).

SUBGOAL 1.1	Assessments obtain a comprehensive view of at least four of the following domains of children's learning, development, and competencies.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
1.1.1	Assessments capture culturally and linguistically relevant domains of development, learning, and competencies and can be adapted to capture additional areas of development as needed.	— Assessments measure key domains of learning, development and competencies that are culturally and linguistically appropriate and elevated by parents/guardians/caregivers, early educators, and Advisory Group members with expertise in equity–informed and culturally responsive literature and research methods with prioritized populations.[11] — Assessments capture domains of learning that reflect key strengths for each of the prioritized populations, as defined in the existing research literature and further endorsed by parents/guardians/caregivers, early educators, and Advisory Group members. — Assessments are flexible enough to allow for additional skills to be captured within targeted domains of language, literacy, math, science, technology, and engineering, executive functioning, approaches to learning, social skills and emotional well–being, and physical development (with additional skills identified and endorsed by panels of parents, early educators, early childhood stakeholders and policymakers, and Advisory Group members).
1.1.2	Assessments capture oral and receptive language competencies.	— Assessments align with early learning standards and capture indicators of children's skills within the developmental domain of language as follows: oral narrative/storytelling (e.g., telling a new or unique account or story with minimal adult prompts), receptive and expressive language, vocabulary, and code switching (particularly for children who speak English and Spanish or different dialects of English).
1.1.3	Assessments capture literacy competencies.	— Assessments align with early learning standards and capture indicators of children's skills within the developmental domain of literacy as follows: listening comprehension, decoding, alphabetic and word knowledge, phonological awareness, concepts about print (including book knowledge), writing letters and words, and writing stories.
1.1.4	Assessments capture a broad range of math competencies.	— Assessments align with early learning standards and capture indicators of children's skills within the developmental domain of math as follows: number (counting, subitizing, comparing numbers), operations (composition of numbers and arithmetic, patterns, and algebraic thinking), geometry and spatial sense (describing, analyzing, comparing, and composing 2D and 3D shapes, spatial visualization, spatial orientation), measurement (length, area, volume), and classification and data (creating and comparing graphs).

[11] See page 35 for Advisory Group list. Additional advisors may be added in later phases of the Measures for Early Success Initiative.

1.1.5	Assessments capture science, technology, and engineering (STE) competencies.	— Assessments align with early learning standards and capture indicators of children's skills within the developmental domains of science, technology, and engineering as follows: prediction, symbolic mapping, relational reasoning, spatial reasoning, and independent inquiry.
1.1.6	Assessments capture executive functioning competencies.	— Assessments align with early learning standards and capture indicators of children's skills within the developmental domain of executive functioning as follows: inhibitory control, working memory, cognitive flexibility/attention shifting, and attention focus/attention control.
1.1.7	Assessments capture approaches to learning.	— Assessments capture indicators of children's skills within the developmental domain of approaches to learning as follows: critical thinking, creativity, engagement, initiative, planning, problem solving, curiosity, agency, and enjoyment.
1.1.8	Assessments capture social skills and emotional well–being.	— Assessments capture indicators of children's skills within the developmental domain of social skills and emotional well–being as follows: emotional and behavioral regulation, emotion knowledge, emotional expressivity, cooperation/teamwork, conflict resolution, confidence (including in learning new skills), self–worth, self–concept, social identity, empathy, behavioral code switching (as somewhat distinct from code switching for children who speak multiple languages), resilience in the face of challenges, and coping.
1.1.9	Assessments capture physical development.	— Assessments capture indicators of children's skills within the developmental domain of physical development as follows: gross and fine motor skills.[12]

SUBGOAL 1.2	Assessments are dynamic and adaptive with children's learning, development, competencies, and age.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
1.2.1	Assessments incorporate age–appropriate, culturally responsive items for 3– to 5–year–old children that vary with age of child and development in domain–specific areas.	— Assessments are adaptive to where children are in their learning, development, and competencies (e.g., any scoring rules, such as setting basal, ceiling, routing, or stop rules, are automated in the administration).

[12] Gross and fine motor skills are often assessed in other settings that support young children and families, such as pediatric settings. Such skills are unlikely to be an area that is heavily focused on for assessment solutions and innovations as part of the Measures for Early Success Initiative.

— Assessments leverage prior information about children—and their prior performance on assessments—to minimize repetition of items for skills that children have mastered when assessed at different time points over the program year.

— Assessment items and stimuli are responsive and adaptive to children's cultural backgrounds and identities as agreed on by early educators and parents of young children within the priority population groups (> 80% across all items; > 70% for individual items).

— Advisory Group members, early educators, and parents of young children agree (> 80% across all items; > 70% for individual items) that items are developmentally appropriate within child age bands (3–4 years and 4–5 years).

— Domain-specific scores demonstrate growth across three-month testing increments (difference between successive time points, $p < .05$) on average across the study sample. Distribution of continuous scaled scores is normal within age bands ($-.5 <$ skewness $< .5$).

SUBGOAL 1.3	Assessments are culturally and linguistically relevant for 3- to 5-year-old children, including those who are Black or Latine, speak Spanish and English, or experience poverty.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
1.3.1	Assessments capture domains of learning that reflect the cultural and linguistic assets and strengths of 3- to 5-year-old children who are Black or Latine, speak Spanish and English, or experience poverty, with diverse geographic representation across the United States.	— Assessments capture domains of learning and development that reflect the cultural assets and strengths of young children as identified—during the research and development phase and once completed—by: • Panels of families of young children from prioritized populations across the U.S. (Black families, Latine families, Spanish- and English-speaking families, or families experiencing poverty) and representing diverse perspectives for each group;[13] • Panels of experienced Pre-K educators of 3- to 5-year-old children from prioritized populations recruited from across the U.S. and representing diverse perspectives for each group; • A panel of state and local Pre-K and early care and education administrators and oversight agencies over early learning standards across the United States and territories; • A panel of researchers (Advisory Group Equity Experts) with expertise in equity-informed, culturally responsive research methods and research with prioritized populations; and • A panel of researchers (Advisory Group Content Experts) with expertise in specific content domains.

[13] See text in the introduction section on background for choosing these priority populations and an acknowledgment that they do not include all historically marginalized groups in the U.S.

1.3.2	Target constructs captured with assessments demonstrate content validity and conceptual equivalence when examined within and across children from four prioritized populations with diverse geographic representation across the United States.	— The meaning, relevance, and interpretation of specific items that capture target, underlying constructs are comparable on average during two rounds of cognitive testing with panels of: • Children from the prioritized populations; • Parents/guardians of children from prioritized population backgrounds (including parents of children who are emergent bilinguals who speak Spanish and English); and • Early educators who are members of and/or work directly with children with prioritized population backgrounds. — An independent review of items and assessment stimuli by a panel of geographically diverse experts, experts in culturally responsive assessments, and experts on emergent bilinguals learning Spanish and English (subgroup of Advisory Group members), educators, and families agree that measured learning domains, items, and stimuli are: 1) culturally and linguistically appropriate; and 2) inclusive of varied dialects with: • > 80% of overall agreement across items and/or stimuli for each target learning domain. • > 70% on each individual item or tested stimuli. — Spanish and English components of assessments clearly reflect activities (as agreed upon by Spanish–speaking parents, educators of young children who speak Spanish and English, and Advisory Group members with expertise in emergent bilingual assessment) done to develop language–specific assessments within the target language (rather than relying on translation of items across languages).
1.3.3	Assessments are administered bilingually to children who speak English and Spanish.	— Assessments allow children to respond in either English or Spanish, meaning that children can switch between languages throughout the assessment activities and receive continued prompts in both languages as needed to ensure their understanding of the content. — Conceptual scoring is incorporated to capture holistic understanding of skills across both English and Spanish languages.
1.3.4	Assessments can generate overall scores for target domains as well as Spanish– and English–specific scores where appropriate.	— For children who speak English and Spanish, assessments can further capture language–specific skills. — Language–specific skills can be compared to criterion–referenced standards. — Language–specific skills can be used to examine individual children's growth within English and Spanish.

SUBGOAL 1.4	Assessments capture children's skills in objective, strengths–based ways.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
1.4.1	Assessments capture measures of children's learning across target domains that minimize reporter bias.	— Assessments primarily rely on direct assessments to capture children's learning, development, and competencies. — Assessments can provide opportunities for educators to report on children's development as a supplement to direct assessment information. — Assessors (typically educators) are able to provide contextual/situation–specific information about the assessment session(s) but do not make ratings of children's skills.
1.4.2	Information captured by artificial intelligence processors or technology reflect minimal cultural, linguistic, and geographic biases.	— Any information captured or scored by assessments using artificial intelligence processors or technology (such as natural language processing systems) has established empirical evidence demonstrating that it is reliable and not biased against children from the priority populations, specifically Black or Latine children, children who speak Spanish and English, or children who experience poverty, with diverse geographic representation across the United States.
1.4.3	Information minimizes cultural or linguistic biases.	— Minimal (< 5%) Differential Item Functioning (DIF) for prioritized groups of children relative to comparison peers and across diverse settings (e.g., Head Start, private and public community–based, public school). — Continuous scale assessment scores demonstrate normal distribution (skewness range is between –.5 and .5) for children within and across prioritized populations and diverse settings. — Domain–specific standardized indicators of reliability and validity (see Goal 2) are consistent within and across prioritized populations and diverse settings. — Advisory Group members agree (> 80%) that items are captured with minimal cultural or linguistic bias.
1.4.4	Information is captured without temporal–, context–, and situation–specific biases.	— Test–retest reliability > 80% when same children are tested within 2 weeks in a different setting or situation (e.g., privately in the hallway vs. during small groups). — Indicators of reliability and DIF are consistent (defined as no more than 10% difference) when assessments are collected in the fall, winter, and spring of the academic years.

GOAL

2

Psychometrics

Instrument(s) collect objective information to produce psychometrically sound and valid data that reflects minimal statistical bias.

Moving towards a more equitable framework for psychometric validation, assessment data will be accurate and relevant for all children. The degree to which this goal is met can be evaluated against criterion–referenced standards to better understand developmental growth trajectories for individuals and groups of children, as well as create greater alignment with state benchmarks for learning. This data can support a data pipeline from kindergarten to 3rd grade to track children's developmental growth and learning over more expansive stretches of time, which will provide rich data for the early childhood education field and continue to inform the development of assessments. Last, but most important, it is expected that assessment items will create comparable and equal results for children irrespective of their race, ethnicity, income, various early child care settings (e.g., public schools, community–based organizations, Head Start centers, etc.), or geographic areas within the United States.

SUBGOAL 2.1	Assessments generate valid, psychometrically sound, and useful information for multiple purposes.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
2.1.1	Assessments generate comparable construct–specific scores—with high levels of content validity as described in prior goals—across groups of 3–, 4–, and 5–year–olds.	— Assessments capture growth relative to a criterion (i.e., what children know and are able to do) developed specifically for priority groups with a representative sample of 3–, 4–, and 5–year–old children from diverse settings and geographic regions of the United States. — Criterion–referenced standards are available for each domain of learning and competency within–age for children ages 3, 4, and 5. — Domain scores can be compared across ages to examine growth relative to criterion–referenced standards. — Assessments yield reliable and valid scores within each age group (3, 4, 5) as defined below.
2.1.2	Assessments generate stable, reliable, and continuous domain–specific scores.	— Assessments demonstrate test/retest reliability ≥ .80 when administered within 2 weeks of each other. — Assessments generate continuous scale scores for each child within learning domains and also provide confidence intervals that support interpretation and allow practitioners to quantify uncertainty in scores. — When collected in a racially and socioeconomically diverse sample of children, within–domain scores have a normal distribution across the sample (e.g., –.5 < skewness < .5) and within prioritized populations. — Assessment scoring methods are valid for intended use case and purpose (e.g., using item response theory to capture unidimensional constructs). — Information from assessments can be continually reviewed to ensure that constructs remain reliable over time across prioritized populations and updated appropriately if needed.
2.1.3	Assessments demonstrate concurrent validity.	— Assessments demonstrate evidence of: • Convergent validity: scores within domains are correlated at .50 or higher on conceptually similar, reliable, and valid assessment tools capturing similar domain(s) (recognizing that existing measures face challenges with statistical bias and elevating the strengths of children in the priority groups).

		• Divergent validity: scores within domain are weakly correlated (.40 or less) with measures capturing conceptually distinct skills. — Information from assessments can be continually reviewed to ensure that constructs remain valid over time across prioritized populations and updated appropriately if needed.
2.1.4	Assessments capture growth over time within domains for individual children.	— Continuous scale scores (for domains where growth is expected during the Pre–K year) demonstrate growth across three–month testing increments (difference between successive time points $p < .05$) on average across study sample and within priority groups. — Changes in scores over time for individual children are larger in magnitude than measurement order (e.g., on a metric such as the conditional standard error of measurement).
2.1.5	Assessments demonstrate evidence of measurement invariance across demographic groups.	— When comparing groups of children by race/ethnicity, home language (including monolingual English–speaking children, children who speak English and Spanish, children who speak African–American English (AAE)), and family income (children experiencing poverty and those not experiencing poverty), assessments demonstrate evidence for measurement invariance ($\Delta CFI < -.01$) when directly evaluated in a confirmatory factor analysis framework (standards summarized in Putnick & Bornstein, 2016). — Limited evidence of differential item functioning (DIF) across items/stimuli (< 5%) when evaluating DIF with a purification process (i.e., removing items one at a time to purify the total conditioning score) [limited to examining constructs where DIF comparison is appropriate].
2.1.6	Assessments demonstrate longer–term predictive validity for schooling, academic, and social–emotional outcomes in kindergarten and 3rd grade.	— Domains captured at end of year are associated with state kindergarten entry (and end–of–year) scores as well as 3rd grade state tests and other district/state measures captured in kindergarten to 3rd grade in that domain, adjusting for variation in the sample's demographic characteristics and other potential confounding variables. — Domains captured at end of year are associated with measures of social–emotional well–being and behaviors captured in kindergarten entry assessments and by elementary schools in the later grades.

SUBGOAL 2.2	Assessments generate comparable and equivalent information for target constructs, across children from different racial, ethnic, linguistic, and socioeconomic backgrounds, and diverse geographic locations across the United States.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
2.2.1	Assessments generate comparable and equivalent information for target constructs, regardless of racial, ethnic, and linguistic backgrounds, or experiences with poverty, across different settings and geographic areas across the United States.	— Minimal DIF across items (< 5%) — for assessments where DIF analysis is appropriate and using a purification process (see above) when comparing children with similar domain–specific total scores who: 1) are monolingual English speakers, speak English and Spanish, and speak AAE; 2) self–identify as Black, Latine, Asian, white, or are another race/are multi–racial; 3) do or do not experience poverty; and 4) are from different settings and geographic areas across the United States. — Assessments demonstrate construct scores with similar means, standard deviations, minimums, and maximums across priority groups when comparing Black and Latine children with white children, children experiencing poverty with those not experiencing poverty, and monolingual English speakers with children who speak English and Spanish and those who speak AAE, representing a geographically diverse population of children (< .05 SD difference across groups; $p > .10$). — Assessments leveraging machine learning demonstrate statistical parity scores within groups in the range of .90—1.1 with minimal differences (maximum .05) between priority populations and reference groups.

SUBGOAL 2.3	Assessments generate objective information about children's skills within and across time.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
2.3.1	Assessments capture information with minimal amounts of rater bias.	— Assessments primarily rely on direct assessments to capture children's learning, development, and competencies. — Assessments can provide opportunities for educators to report on children's development in addition to direct assessment information.
2.3.2	Information captured by artificial intelligence processors or technology are culturally, linguistically, and geographically unbiased.	— Any information captured or scored by assessments using artificial intelligence processors or technology (such as natural language processing systems) has established empirical evidence demonstrating that it is unbiased and reliable for all children of Pre–K age, including Black and Latine children, children who speak Spanish and English, children who experience poverty, and children who are geographically diverse across the United States.

GOAL
3

Experience

Instrument(s) are enjoyable and engaging for children and easy for educators to collect.

Young learners need fun, engaging assessment experiences that are appropriate for a range of diverse learners. Assessments should be brief and developmentally appropriate for 3– to 5–year–old children.

Educators must also feel supported in their efforts to conduct assessments and be provided data that addresses their needs, rather than simply feel it is time spent away from valuable learning opportunities in Pre–K. To this end, the setup, administration, and data collection process related to the assessment should feel intuitive, be closely integrated with Pre–K activities, and be viewed by educators as helpful for students to reach their learning goals. Training for educators on the assessments should also be intuitive, not burdensome, and account for the broad spectrum of technological proficiency among the teaching workforce. These methods for data collection can be balanced more heavily on direct assessment where possible to increase reliability and validity and triangulated with insights from educator and family observations.

Families will also need to be engaged as important consumers of assessment data and findings. As a starting point families should first be aware of when their children are assessed, the purpose of the assessment, and which populations of children and families were engaged in the development of the assessment tool. All written communications will be provided with at least English and Spanish as primary languages.

SUBGOAL 3.1	Assessments are engaging and can be self–guided by children.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
3.1.1	Assessments are play–based.	— Assessments are interactive and interesting to young children with engaging scenarios, graphics, and sounds that are culturally relevant.
3.1.2	Children easily learn how to complete the assessments.	— Assessment rules of administration are consistent across assessments and have self–guided prompts to orient children. — Introduction, prompts, and practice items teaching children how to complete the assessment are presented in a brief, straightforward, and engaging way. — Over 75% of children demonstrate comprehension of the assessment (e.g., successfully learn and pass any introduction and practice items on the first attempt).
3.1.3	Assessments are designed to create auditory, visual, and tactile experiences that are developmentally appropriate for 3– to 5–year–old children.	— Assessment digital interfaces meet accessibility guidelines, WCAG 2.0 and 2.1, set by W3C. — Color scheme used accounts for color blindness and other potential visual issues. — Stimuli are presented in large graphics. — Touchscreen experience is simple and straightforward (e.g., does not rely on dragging and dropping objects, which can be difficult for young children developing fine motor skills).
3.1.4	Assessments are brief and age–appropriate in length to administer and for children to complete.	— The time per administration per learning domain is less than 10 minutes, on average. — A sample of 3– to 5–year–old children from a diverse range of settings do not appear to be fatigued when engaging with the assessments (according to a researcher observer and the child's educator).
3.1.5	Children find the assessments to be joyful and fun.	— A sample of 3– to 5–year–old children from a diverse range of settings appear to enjoy and have fun engaging with the assessments (according to a researcher observer and the child's educator). — Assessments are positively rated by a sample of 3– to 5–year–old children from a diverse range of settings: • Self–report (75% or higher) that the assessments were fun to do, that they like them, and want to do them again. • Self–report (25% or lower) that the assessments were boring.

3.1.6	Assessments are technology-enabled and do not require external support or stimuli for children to complete.	— Assessments are technology-enabled, are self-guided, and can be completed by children with minimal support from educators in everyday Pre-K settings. — Over 90% of children can successfully and independently complete the assessments in their entirety.

SUBGOAL 3.2	Assessments can be integrated into everyday classroom activities seamlessly.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
3.2.1	Administration of the assessments can be embedded within typical Pre-K routines and does not take away from other activities.	— Time spent in typical instructional activities is largely unchanged (or potentially increased) before and after the take-up and implementation of the assessments in diverse Pre-K settings. — Assessments are designed to be used in a variety of activities throughout the day (e.g., individual choice time or project-based time).

SUBGOAL 3.3	Educators find assessments are easy, brief, and intuitive to administer.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
3.3.1	Setup of assessments is brief and intuitive and is not burdensome.	— Assessments require, on average, less than 2 minutes for an educator or assistant educator to set up per administration per child without assistance. — Assessments require minimal training or have a self-directed training session or supports that takes less than 10 minutes for an educator or assistant educator to review. — 90% of educators report that rules, guidelines, and login for interface are not redundant with other systems, are easily recalled, and are implementable. — 90% of educators can set up the platform and assessment (e.g., login, sign up, children in system) without assistance. — 90% of Pre-K policymakers and center directors agree that required training on assessments is feasible to implement and requires minimal staff time or program resources. — Educators' instructions for setup are in English but can toggle to Spanish for educators who may be more proficient in Spanish.

		— The majority of a sample (e.g., 75% of sample) of educators and assistant educators who have used the assessments for at least 6 months report that they *do not* perceive the setup and administration of the assessments to be more burdensome than other assessments.
3.3.2	Educator training and guidance materials about assessments provide easily understandable information about the purpose of the assessments, and how the information being collected will be used.	— 90% of educators report feeling comfortable describing the purpose of the assessments and how the information being collected will be used. — 90% of educators can accurately describe the purpose of the assessments and how the information being collected will be used to other educators and to parents/guardians/caregivers.

SUBGOAL **3.4**	Assessments are designed for comparable administration and scoring in both English and Spanish.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
3.4.1	Assessments are developed for comparable administration in both English and Spanish.	— Administration experience of assessments is designed to be similar for children who are monolingual English, monolingual Spanish, and emergent bilinguals who speak Spanish and English. — Assessments are designed to switch between English and Spanish based on emergent bilinguals' language ability and comprehension. — Assessments are designed so emergent bilinguals can respond to item prompts in either English or Spanish.
3.4.2	Assessments allow for easy identification and accurate routing, and administration in Spanish or English.	— Assessments include an integrated, brief (< 5 min), and child–guided approach to determining the relevant language(s) in which the assessments for each learning domain are administered in English or Spanish. — Assessments integrate opportunities for educators or parents/guardians/caregivers to report on children's languages spoken other than English. — Assessments can easily toggle back and forth to change the language of the assessments between English to Spanish, as needed. — A panel of educators, families, and researchers agree (> 80%) that the assessments' approaches to determining the relevant language(s) of administration are culturally and linguistically appropriate and do not detract from the enjoyability of the assessments.

GOAL

4

Usefulness

Instrument(s) generate timely, easily accessible, readily digestible, and understandable information for several purposes.

The same suite of tools or a single tool should serve priority functions for educators, families, and system administrators. That is, a set of assessment tools should: (a) provide real–time information to help educators understand children's learning and tailor instruction appropriately, (b) help document all children's progress, (c) eliminate redundancies in data collection, (d) facilitate communication among educators, families, and other stakeholders supporting children's learning and development, and (e) increase understanding of how children's data are used to inform program and policy decisions. To support respectful, bidirectional, and accurate communication, frameworks for interpreting and delivering culturally responsive insights from data will be needed.

SUBGOAL 4.1	Assessments regularly generate meaningful and actionable information about children's learning, development, and competencies in separate early learning domains for multiple purposes.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
4.1.1	Assessments produce results that can be used to identify how children are learning and tailor instruction to support children's development.	— Assessments produce results for each child at least 6 times—or as frequently as needed by the educator to support an individual child's development—during the program year that: • Can produce point-in-time holistic profiles for child development across multiple domains. • Can produce reports on individual children's growth and areas for supported learning in domain-specific areas from one assessment period to the next, from the beginning of the year to the most recent assessment, and from the beginning to the end of the program year. • Can produce reports on individual children's performance relative to overall classroom/group performance. • Can suggest groupings of children with like abilities or mixed abilities in small groups. • Can produce reports on overall classroom/group performance across multiple domains.
4.1.2	Assessments produce results to document children's progress compared to criterion-referenced standards.	— Assessments produce results for each child at least 3 times during the program year that: • Can be used to measure children's development and growth at beginning and end of the program year against a criterion (e.g., a learning objective that has age-normed standards) designed specifically for the priority groups. • Can be used to inform children's strengths, areas for growth, and needs, so that transitions, supports, and instruction can be aligned in the next program year/grade.
4.1.3	Assessments produce results for program- and system-level planning and improvement purposes.	— Assessments produce results for programmatic and system purposes at least 2 times per program year that: • Can produce program-level estimates about children's strengths and areas for additional support for the current program year for a given program. • Can produce comparative program-level and system-level estimates about children's strengths and areas for additional support for prior program years to look at trends for a given program. • Can produce population-level estimates about children's strengths and areas for additional support for the current program year within a system. • Can produce comparative population-level estimates about children's strengths and areas for additional support for prior program years to look at trends in local populations and 3- to 5-year-old children served by a particular system.

		• Allow for disaggregation of population–level estimates about children's strengths and areas for additional support for the current program year and prior program years for subsets of the population of children within a system by different program type, for different communities, and for children who are Black or Latine, who are emergent bilinguals who speak Spanish and English, or who experience poverty.
4.1.4	Results from the assessments are made available in a timely fashion for different purposes.	— Results are available to educators and families within 24 hours of a child's completion of an assessment. — Results from the assessments for programmatic and system purposes are available to intended end–users within 24 hours of completion of assessments for all children in the program or system. — The majority of a panel of educators, families, administrators, and policymakers who have used the assessments for at least 6 months report that they obtain and review timely and relevant information about children's results in line with their prioritized interests and intended purposes when compared with other assessments.
4.1.5	Assessments allow educators and parents/guardians/primary caregivers to share insights about the assessment information to enhance its usefulness and relevance for different end–user perspectives.	— Assessments provide educators and parents/guardians/primary caregivers opportunities to share their user feedback, which can be applied to guide updates and adaptations to the design and implementation of the assessments. — Assessments allow educators to provide input on children's development in different domains of learning. — Assessments allow parents/guardians/primary caregivers to provide input on children's development in key domains of learning they present outside of the Pre–K context.

SUBGOAL 4.2	Assessment results are easily accessible and useful for families.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
4.2.1	Parents/guardians/caregivers are able to easily access and understand insights about children that are communicated in a culturally responsive way.	— Assessments include an online portal with a dashboard that presents assessment results in a user–friendly, visually appealing way in English and Spanish. — Assessments produce reports that explain what concepts are measured, why their children are learning them (learning objectives), which topics their children have mastered, and where their children are still learning content, skills, and competencies.

		— Dashboards for the assessments offer filters and choices about what comparisons parents may want to see (e.g., own child's growth and development over time vs. own child compared to criterion-referenced standards at a single point in time). — Dashboards for the assessments are multi-platform accessible (e.g., mobile-friendly-viewable on a mobile device or on computer/laptop). — The majority of a panel of families (e.g., 75% of sample) who have used the assessments for at least 6 months report that it is easy to access and to understand information about children's learning progress and that the information is communicated in a culturally responsive way.
4.2.2	Parents/guardians/primary caregivers are able to regularly see their child's progress.	— Assessments can provide parents/guardians/caregivers updated, clear information on each child's strengths and key areas for growth for different domains of learning, development, competencies via dashboard/online system: • Over the course of the program year and at end of year (at least 4 times per year). • In relation to standard or benchmark (at least 3 times per year).
4.2.3	Assessments offer clear use-case information to enhance the relevance, interpretability, and understanding of assessment information for parents/guardians/primary caregivers.	— Assessments or dashboards for assessments provide parents/guardians/primary caregivers with clear, timely, and understandable information in English and Spanish about when children will be assessed and how the information will be used by Pre-K educators, program administrators, and systems. — The majority of a panel of families who have used the assessments for at least 6 months report that the assessment information is meaningful to them, they know when their children will be assessed and what the assessments will cover, and they can describe how the information will be used by Pre-K educators, program administrators, and systems to support children's early learning and development. — Assessment information is provided to parents/guardians/caregivers that includes easily understandable information that explains how the assessments were developed and tested with linguistically and culturally diverse samples of children.

SUBGOAL 4.3	Assessment results are easy to use and useful for educators.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
4.3.1	Educators are able to easily access information about each child's and entire classroom's or group's learning progress.[14]	— Assessments include an online portal with dashboards that presents results in a user–friendly, visually appealing way. — Dashboards for the assessments offer filters and choices about what insights educators may want to see (e.g., each child over time vs. each child compared with the entire classroom/group or criterion–referenced standards at a single point in time). — Dashboards for the assessments are multi–platform accessible (e.g., mobile–friendly: viewable on a mobile device or on a computer/laptop). — 90% of a panel of educators who have used the assessments for at least 6 months report that it is easy to access and to understand information about each child and the entire classroom's or group's learning progress, and use the information in actionable ways (e.g., to plan instructional activities, facilitate communication with families, etc.).
4.3.2	Educators are able to regularly see each child's and entire classroom's or group's learning progress.	— Assessments can provide educators updated, clear information on all children's strengths and key areas for growth for different domains of learning, development, competencies in the classroom/group via dashboard/online system: • Over the course of the program year and at end of year (at least 4 times per year). • In relation to standards or benchmarks (at least 3 times per year).
4.3.3	Assessments allow educators to regularly use information to differentiate instruction for children based on children's abilities and areas for development and mastery.	— Dashboards for the assessments can help group children (e.g., based on language and literacy assessments). — Dashboards for the assessments can generate suggestions each week for what educators can focus on for key groups of children that week/month. — Pre–assessment and during–assessment information can help identify where children may need more practice or where educators need to use more culturally appropriate language to promote understanding.

[14] In mixed delivery systems, Pre–K children may be divided into classrooms or smaller groups within one space.

4.3.4	Assessment results are easy to access with other management tools used to plan, organize, and monitor classroom/group and child progress.	— Educators can easily import classroom/group rosters into a dashboard at the start of the year and update rosters as necessary. — Results from the assessments can be easily exported and merged with classroom/group rosters, "gradebooks," and other management information systems. — Data, scores, and other information about children are clearly owned by the Pre–K program using the system, and can be fully and easily downloaded in machine–readable export formats, such as .xml, .json, or .csv.
4.3.5	Assessments offer clear interpretability of information.	— Dashboards for assessments provide online visualizations of final scores (e.g., bar charts and/or color codes) over words (i.e., they avoid terms like "proficient" that have normative judgment). — Dashboards for assessments produce reports that provide prompts/explanations of concepts measured, and updated, clear information on each child's—and all children in the classroom's/group's—strengths, how children are progressing relative to early learning standards, and key areas for growth for different domains of learning, development, and competencies.

SUBGOAL 4.4	Assessments generate information that can facilitate communication among educators, families, and other stakeholders supporting children's development.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
4.4.1	Assessments allow for bidirectional communication between parents and educators about children's learning and progress, and allow parents and educators to share information with other stakeholders who also support children's learning and development.	— Assessments include a dashboard or online portal that allows for real–time communication about children's developmental progress as educators and/or parents view results. — Assessments offer a dashboard or online portal with filters and choices about what comparisons other stakeholders may be interested in (e.g., aggregated or disaggregated data) to produce reports that can be shared via password–protected links.

GOAL
5

Scalability

Instrument(s) can be administered at scale in publicly funded Pre–K systems.

Reimagining the existing assessment and data landscape about young children's early learning in Pre–K systems on a large scale will likely require leveraging innovative solutions and their technical viability to address challenges of the current context. Emerging technologies may act as a critical lever for scaling and lowering the costs of collecting assessments and enhancing the accessibility of the resulting information to inform data–driven insights. As we begin to realize the potential of widespread collection of assessment tools on a regular basis that can be supported by technology–based innovations and data systems, developers of future–state assessments must also take into account accompanying issues around protecting student privacy, rules around the interoperability of systems, and regulations that may act as bottlenecks or as enablers to the flow of data. In addition, assessment results will need to be calibrated with existing early learning standards and outcome frameworks to be relevant on a large scale, given the diversity of federal, state, and local policies, regulations, and oversight agencies that make up Pre–K systems nationally.

Care must also be taken to address resource limitations that Pre–K programs may face in securing different types of technology or even access to Wi–Fi. Even in cases where data can be entered in offline mode, challenges with uploading the data online may prove taxing. A careful consideration of these contextual issues will increase the likelihood that any future assessment solutions will be adopted and utilized by states, educators, and families.

SUBGOAL 5.1	Assessments are affordable for publicly funded Pre-K systems and centers to administer. (Feasible price and time burden target points are currently being determined through discussions with Pre-K system leaders, program administrators, and educators.)

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
5.1.1	There are low costs and burdens to adopt the assessments for Pre-K systems and programs.	— Cost of initial take-up is reasonable and feasible as agreed on by a panel of program administrators, center directors, and policymakers (costs here include the hardware and software costs to start up, and staff time to learn a new system of assessments, such as training time for educators and administrators, educators' and administrators' time to review and interpret data, and costs for IT support staff to launch, divided by the total number of children in a program or system). — Families, educators, and administrators in diverse early learning settings perceive the benefits of taking up and collecting the assessments to outweigh the costs of doing so after having used the assessments for at least 6 months. — Families, educators, and administrators are able to understand information from reports quickly and efficiently. Panel of families, educators, and administrators agree (> 80%) that implementing assessments does not detract from time spent with children or typical learning activities.
5.1.2	The infrastructure and hardware equipment required to administer the assessments are specified to the typical IT infrastructures of Pre-K programs and systems.	— Assessments and dashboards for assessments are compatible with Windows, Mac, Chrome OS, iOS, and Android operating systems. — Implementing the assessments in Pre-K settings requires minimal new investment in technology (e.g., no more than two iPads, Microsoft Surface Pros, or Chromebooks [no more than three generations behind the latest version] per classroom; two headphones/microphones). — Implementation of the assessment system should require no more than 1.25 Mbps Wi-Fi internet connection per child with a total minimum of 10 Mbps per site and connectivity quality of service equivalent to fiber or cable connections (e.g., no GEO satellites or 4/5G wireless connections). — Accessing and obtaining results and summary reports of information collected by the assessments or the dashboards for assessments require no more than a laptop/computer with available Wi-Fi to be accessible by the end-user. — Assessments can be administered when Wi-Fi is available and can also be conducted asynchronously in off-line mode, as needed, with data capture when Wi-Fi is unavailable. — Assessment platforms can pre-download assessments in advance of use, to ensure that assessments can be administered without reliable network connectivity.

		— Assessment results can be uploaded when network connectivity becomes available and are reliable and persistently stored on the device until they are uploaded. — Multiple students can take an assessment on a device without a network connection, and all their results can be uploaded when the network is restored. — Information from assessments can be accessed immediately with internet access or asynchronously when internet access becomes available. — A panel of Pre-K center directors, educators, and local policymakers agree that implementation of assessments will face limited challenges due to internet access.
5.1.3	Sustaining administration of the assessments is affordable within available public funding for Pre-K systems and is not burdensome from an end-user perspective.	— Assessments can be easily and readily updated with minimal end-user involvement (e.g., receive automated software updates). — Assessments can be used across a diverse range of Pre-K settings and with varied curricula. — Cost of continued administration after initial take-up is reasonable and feasible as agreed on by a panel of program administrators, center directors, and policymakers (costs here include the time for educators/assistant educators to administer assessments—e.g., 5 minutes per week per child).
5.1.4	Assessments generate information that can be integrated into existing systems or are stand-alone with low cost and burden to Pre-K systems and end-users.	— Data systems that organize assessment information are adaptable and flexible enough to coordinate with existing data platforms (e.g., a classroom rostering system used by the Pre-K program). — Assessment systems are compatible with widely used operating system supporting alternative data platforms. — Data systems for assessments can also exist as stand-alone platforms (depending on school/center preference).

SUBGOAL 5.2	Assessments collect and store information in a private and secure way.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
5.2.1	Assessments adopt technical parameters and processes aimed at ensuring the privacy and security of information collected and stored about children and educators/ assistant educators.	— Assessments provide information about how the data collected about children is maintained and secured to educators and parents/guardians/caregivers. — Assessments maintain and store data about children in compliance with requirements of Family Education Rights & Privacy Act (FERPA), the Children' Online Privacy Protection Act (COPPA), and the Children's Internet Protection Act (CIPA).

SUBGOAL 5.3	Assessments can be regularly, easily, and reliably administered and completed as intended by design.	
ID	**CRITERIA**	**ASPIRATIONAL TARGET THRESHOLDS**
5.3.1	Assessments offer educators implementation supports and quality controls to ease planning and collection of assessments.	— Assessments offer clear timelines, reminders, and quality controls to support educators in planning for and administering assessments with each child. — Educators reliably capture assessments for each child with minimal errors for at least 6 months of use. — Educators report that assessments facilitate their planning, administration, and collection of the assessments with each child without being burdensome. — Assessments can be administered and completed with fidelity in line with the intended administration protocols for over 95% of a sample of 3– to 5–year–old children who are enrolled in Pre–K settings across the United States (which may include school–based, community–based, home–based, and Head Start). — Any administration and scoring rules for the assessments (e.g., setting basal, ceiling, routing, and/or stop rules) are automated and programmed directly into the assessments to ensure consistency across administration and scoring.

SUBGOAL 5.4	Assessments can be updated to enhance the sustainability and relevance of the information collected.

ID	CRITERIA	ASPIRATIONAL TARGET THRESHOLDS
5.4.1	Assessments allow for easy integration of new assessments of emerging domains of children's learning, development, and competencies, and replacement or adaptation of existing assessments in response to insights gained from end–user feedback or administration, and measurement developments in the field.	— Assessments are designed with flexible in–app update capability (defined as: updates can be downloaded in the background; after downloading, assessors are asked to restart the assessments; and during the restart period, updates are installed and then automatically restarted). — Assessments can be regularly updated and adjustments to existing assessments or new assessments can be added without loss of collected information about children's learning, development, and competencies and without compromising the educators' and children's use experiences.

Acknowledgments

We are deeply grateful for the families, educators, coaches, administrators, and early childhood stakeholders from over 20 states and localities who advised us. We are thankful for the time they took to openly share their hopes and dreams for young children, their experiences and perspectives with assessments, and their lived expertise more broadly. The importance of continuing to integrate feedback and engage in co-design activities with these priority stakeholders, whose voices are often not heard, is a central theme throughout this document.

Additionally, an advisory group was convened by the MDRC and Substantial teams at multiple points during our writing process (see names below). We appreciate their time and willingness to provide their insights about how assessments could better serve the needs of Pre-K educators, families, and young children. Blending policy and research considerations is never simple, but their openness to explore both the complexities of this space, as well as potential solutions, greatly informed the writing of this document. It should be noted, however, that this document does not necessarily reflect their practices or points of view.

The following people synthesized the feedback and authored the narrative and technical specifications of this document: JoAnn Hsueh, Ximena Portilla, Meghan McCormick, and Rekha Balu (MDRC), and Behnosh Najafi (Substantial). Sharon Huang (MDRC), Sheryl Cababa, and Amanda Di Dio (Substantial), and Elizabeth Mokyr Horner, Marquita Davis, and Sara Allan (Bill & Melinda Gates Foundation) reviewed drafts and provided valuable feedback. Joshua Malbin (MDRC) and Deborah Beckwin (Substantial) carefully edited the final draft. Christen Miller (Substantial) designed the layout and prepared the document for publication.

Advisory Group

Alexandra Figueras–Daniel
National Institute for Early Education Research

Alisha Wackerle–Hollman
University of Minnesota

Andres Bustamante
University of California, Irvine

Anna Johnson
Child Trends

Benjamin Shear
University of Colorado Boulder

Bobbi Shannon
YMCA of Central Ohio

Bridget Hamre
Teachstone

Carol Hammer
Teachers College, Columbia University

Catherine Snow
Harvard Graduate School of Education

Chavaughn Brown
AppleTree Institute for Education Innovations

Christina Weiland
University of Michigan

Dana McCoy
Harvard Graduate School of Education

Doug Clements
University of Denver

Emily Hanno
formerly at Harvard Graduate School of Education

Esther Calzada
The University of Texas at Austin

Gigliana Melzi
New York University

Ginny Vitiello
University of Virginia

Herbert Turner
Analytica

Hirokazu Yoshikawa
New York University

Iheoma Iruka
University of North Carolina at Chapel Hill

James Soland
University of Virginia

Jason Sachs
Boston Public Schools

Jelena Obradović
Stanford Graduate School of Education

Jennifer Keys Adair
The University of Texas at Austin

Jennifer Randall
University of Massachusetts, Amherst

Jessica Whittaker
University of Virginia

Julie Sarama
University of Denver

Katherine Zinsser
University of Illinois at Chicago

Kathy Hirsh–Pasek
Temple University

Laura Bornfreund
New America

Lily Nuñez
formerly at Child360

Megan Kuhfeld
NWEA

Michael Rodriguez
University of Minnesota

Milagros Nores
National Institute for Early Education Research

Monisha Gibson
Maritime Odyssey Preschool

Pamela Morris–Perez
New York University

Rachel Abenavoli
New York University

Rashmita Mistry
University of California, Los Angeles

Roberta Golinkoff
University of Delaware

Sandra Barrueco
The Catholic University of America

Scott McConnell
Renaissance Learning

Stephanie Curenton–Jolly
Boston University

Steve Midgley
Learning Tapestry

Susan Adams
Georgia Department of Early Care and Learning

Tamara Halle
Child Trends

Trish Millines Dziko
Technology Access Foundation

Whitney Whealdon
Learning Tapestry

References

Coll, Cynthia García, Gontran Lamberty, Renee Jenkins, Harriet P. McAdoo, Keith Crnic, Barbara H. Wasik, and Heidie Vázquez García. 1996. "An Integrative Model for the Study of Developmental Competencies in Minority Children." *Child Development* 67, 5: 1,891–1,914.

Gardner-Neblett, Nicole, and Iheoma U. Iruka. 2015. "Oral Narrative Skills: Explaining the Language-Emergent Literacy Link by Race/Ethnicity and SES." *Developmental Psychology* 51, 7: 889–904.

Friedman-Krauss, Allison H., W. Steven Barnett, Karin A. Garver, Katherine S. Hodges, G.G. Weisenfeld, Beth Ann Gardiner, and Tracy Merriman Jost. 2022. *The State of Preschool 2021: State Preschool Yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

Head Start Early Childhood Learning and Knowledge Center. 2019. "Head Start Program Facts: Fiscal Year 2019." Website: https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/no-search/hs-program-fact-sheet-2019.pdf.

Home Grown. 2020. "Home-Based Child Care Fact Sheet." Website: https://homegrownchildcare.org/wp-content/uploads/2020/12/HomeGrown_Child-Care-Fact-Sheet_final.pdf.

Hsueh, JoAnn. 2021. "Challenge and Opportunity: Equitable Pre-K Measures for Early Learning." New York: MDRC.

Melzi, Gigliana, Christine McWayne, and Wendy Ochoa. 2022. "Family Engagement and Latine Children's Early Narrative Skills." *Early Childhood Education Journal* 50: 83–95.

Peña, Elizabeth D., and Tamara G. Halle. 2011. "Assessing Preschool Dual Language Learners: Traveling a Multiforked Road." *Child Development Perspectives* 5, 1: 28–32.

Putnick, Diane L., and Marc H. Bornstein. 2016. "Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research." *Developmental Review* 41: 71–90.

Thank you

Bill & Melinda Gates foundation

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

substantial

Suggested citation: MDRC and Substantial, *User–Informed Principles: Developing Assessments for All Early Learners* (New York: MDRC, 2022).