

5-6-2023

How Can Urban Early Childhood Educators Advocate for Accountability in Alternative Assessment Methods

Stephanie Scierka
Concordia University, St. Paul, stephscierka@gmail.com

Follow this and additional works at: https://digitalcommons.csp.edu/teacher-education_masters



Part of the [Early Childhood Education Commons](#), and the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Scierka, S. (2023). *How Can Urban Early Childhood Educators Advocate for Accountability in Alternative Assessment Methods* (Thesis, Concordia University, St. Paul). Retrieved from https://digitalcommons.csp.edu/teacher-education_masters/82

This Thesis is brought to you for free and open access by the College of Education & Humanities at DigitalCommons@CSP. It has been accepted for inclusion in Graduate Teacher Education by an authorized administrator of DigitalCommons@CSP. For more information, please contact digitalcommons@csp.edu.

How Can Urban Early Childhood Educators Advocate for Accountability in Alternative

Assessment Methods

Stephanie Scierka

Concordia University, Saint Paul

ED 590 Research and Complete Capstone

Instructor: Dr. Kelly Sadlovsky

Second Reader: Kristin Goetz

Author Note

A tremendous amount of love and gratitude to my husband, Andy and daughter, Lauren during this journey to pursue my master's degree. It has taken me away from swim meets and family events, but you have both cheered me on through it all. Thank you from the bottom of my heart. I could not have done this without you. To the professors at Concordia University-St Paul that pushed me outside of my comfort zone and helped me become a better educator in so many ways, thank you so much. And to my kindergarten team, thank you for your love, support, and grace when I was overwhelmed. You stepped in, dusted me off, and kept me going. I love you both more than you will ever know, L.B. & A.N.

Table of Contents

Abstract	6
Chapter One: Introduction	7
Important Definitions for Educational Advocates	9
The Start of Inappropriate Practices	11
<i>Assumptions That Led to Counterproductive Educational Acts</i>	11
Effects of These Counterproductive Educational Acts on Students.....	12
Effects of These Counterproductive Educational Acts on Teachers	13
Chapter Two: Literature Review.....	16
<i>Does Standardized Testing Improve Instruction</i>	17
<i>Method</i>	17
Results and Discussion	19
How Educator Experiences with Mandated Tests Effects Views on Test’s Purpose	20
<i>Methodologies</i>	20
Results/Researcher Discussion	21
Link Between Student’s Approach to Learning, Teacher Ratings, and Academic Success	22
What Constitutes as Developmentally Appropriate Practices (DAP) in Assessing Children	24
How Educators Faced the Accountability Challenge with DAP	24

ACCOUNTABILITY IN ALTERNATIVE ASSESSMENT METHODS	4
Adapting, Augmenting, or Extending Curriculum.....	26
Advocating for Adjustments in Multi Language Learner Assessments in Wales	28
<i>Method</i>	28
<i>Limitations/Discussion</i>	29
Linking Authentic Assessments to Accountability Measures	29
Conclusion	30
Chapter Three: Insights, Application, and Future Studies	32
Insights Gained That Advocate for Accountability via Alternative Assessments	32
<i>How to Identify Developmentally Appropriate Assessments</i>	33
Authentic Forms of Assessment.....	34
<i>Observation-based</i>	34
<i>Curriculum Embedded Performance</i>	34
<i>Work Sampling System (WSS)</i>	35
Suggestions for Future Studies and Practices	35
<i>Importance of Social-Emotional Learning</i>	35
<i>Limit Direct-Assessments to Two to Three Times a Year</i>	36
<i>Limiting Standardized Test Weight on Important Educational Decisions</i>	36
Conclusion	37

References..... 38

Abstract

The introduction of district and educator accountability measures based on student assessment scores has created a move away from developmentally, culturally, and linguistically appropriate practice (Bauml, 2016; Cramer et al., 2018; Frede & Ackerman, 2007; Gullo & Hughes, 2011; Miller & Almon, 2009; NAEYC, 2020; Rushton & Juola-Rushton, 2008). A combination of qualitative, quantitative, mixed-method studies and guidelines from the National Association for the Education of Young Children [NAEYC] (2020) demonstrated a need for developmentally appropriate practices when assessing students. The guiding question focused on how urban early childhood educators could advocate for accountability in alternative assessment methods. The research discussed the effects of standardized testing on students and teachers, observed how different educators faced mandated accountability measures, examined guiding principles of developmentally appropriate practices, and demonstrated ways that educators can use research to advocate for alternative assessment methods that show accountability within urban early childhood classrooms. Add information about mindset importance.

Keywords: accountability, authentic assessment, culturally responsive, developmentally appropriate practice, early childhood education, norm-referenced, pedagogy, standardized test

Chapter One: Introduction

With the ever-changing demographics of the public school system, it is crucial to provide assessments that are not only developmentally appropriate but also culturally and linguistically appropriate (Cramer et al., 2018; Gullo & Hughes, 2011; Miller & Almon, 2009; NAEYC, 2020; Wolk, 2011; Yun et al., 2021). Research has shown that children develop on a continuum unable to be captured within standardized accountability measures (Cramer et al., 2018; Gullo & Hughes, 2011; Im, 2017; NAEYC, 2019, 2020; Yun et al., 2021). The factors that influenced the 2001 No Child Left Behind Act (NCLB) were well-intended but not created by educators that knew child development. The lack of developmental knowledge led to standardized tests that were not culturally, linguistically, or developmentally appropriate for the students assessed (Cramer et al., 2018; Gullo & Hughes, 2011; Miller & Almon, 2009; NAEYC, 2003, 2011, 2019, 2020). Years after the creation of NCLB, schools across the United States have reported significant behavior issues, more student suspensions, children struggling with mental health or trauma, and a rise in educators leaving the teaching profession (Cramer et al., 2018; Bodrova, 2008; Rushton & Juola-Rushton, 2008; Ryan et al., 2017; Hine et al., 2022; Wolk, 2011). The discord between early childhood educators and policymakers did not happen overnight.

From the 1954 *Brown v. Board of Education* ruling on equal educational rights; to President Lyndon B. Johnson's 1965 Elementary and Secondary Education Act (ESEA) providing government funding to the poorest schools (AFT, 2015), there have been efforts made to create a better educational system for students in the United States. When ESEA's funding did not show the improvement that many had hoped for, the National Commission on Excellence in Education (1983) submitted a report titled *A Nation at Risk*. That report led to the birth of standardized testing and the use of data points to measure accountability in schools. The original

ESEA bill has been revised and renamed many times. The most recent revision, titled “Every Student Succeeds Act” (ESSA), was made in 2015 by the Obama Administration. ESSA put an end to No Child Left Behind’s (NCLB) adequate yearly progress (AYP) requirements and gave more flexibility to states on how to implement and measure high standards. Even with more statewide flexibility, accountability measures continued to be judged based on norm-referenced assessments (Cramer et al., 2018; Gullo & Hughes, 2011; NAEYC, 2020; Ryan et al., 2017). Each bill made changes intended to help the education system, but the changes continued to follow the same misguided belief; that educator accountability was measurable by a norm-referenced test (Cramer et al., 2018; Gullo & Hughes, 2011; NAEYC, 2020; Ryan et al., 2017).

The following problems arose with norm-referenced tests utilized for accountability measures: the norms did not match the demographics of diverse classrooms; there was no consideration of prior test-taking skills or technical knowledge; norm-referenced tests are not capable of capturing children’s learning and development on a continuum; standardized tests cannot accurately identify all the ways children can show content mastery with a fixed data point (Cramer et al., 2018; Gullo & Hughes, 2011; Im, 2017; Miller & Almon, 2009; NAEYC, 2003,2011, 2020). The problems listed above discussed the testing format. Next, researchers examined how to measure student growth accurately.

Studies have shown the need for educators to understand children’s development and how to properly assess student growth (Im, 2017; NAEYC, 1995; Rushton & Juola-Rushton, 2011; Blessing, 2019; Yun et al., 2021). The research on how children’s brains develop, the continuum of learning, and the individual ways students can show knowledge, has helped educators accurately record valuable information about student content mastery. “Pressuring children to produce abstract academic tasks ignores all the research and knowledge of child development

that early childhood teachers have at their disposal” (Schoenbeck, 2000, p. 21). There is a need for policymakers to examine more research on child development and how to properly assess growth. NAEYC (1995) researchers stated, “For too long we have enabled educational achievement for the very few” (p.3), and educators must learn how to advocate for all students to ensure this ‘enabling’ does not continue. The following research can provide educators with knowledge on how to advocate for accountability in alternative assessment measures by showing where inappropriate practices started, the unintended consequences of inappropriate accountability measures on students and teachers, ways teachers adjusted pedagogy to meet standards more appropriately, and research-based options for alternative.

Important Definitions for Educational Advocates

Accountability

A documented measurement that monitors educator and program quality based on student’s end of year testing scores (Meisels, 2006; Ryan et al., 2017)

Authentic assessment

The age-appropriate evaluation of student learning by observation of children during natural activities instead of technology-based testing structures. For example, when a child shows the teacher an AB pattern made with blocks during active learning, the teacher marks down that the student understood AB patterns using a running record or checklist (NAEYC, 2020).

Culturally responsive

A pedagogical approach that focuses on bringing students’ family structure, cultural experiences, and lived experiences into the classroom and curriculum. Educators also acknowledge how personal culture and family structure affect individual teaching styles.

Educators use this knowledge to build reciprocal relationships with families to ensure there are home-to-school connections within the learning community (NAEYC, 2020).

Developmentally appropriate practice (DAP)

A research-based approach focused on how children learn, develop, and grow on a continuum that correlates with family and community cultures or structures. Educators use knowledge of child development to build meaningful relationships with students and assess learning in a multitude of ways to reach individual children's needs (NAEYC, 2020).

Early Childhood Education

A profession focused on children's social, emotional, academic, and developmental learning from birth to eight years old (NAEYC, 2020).

Norm-referenced

A predetermined set of outcomes, identities, and actions to measure what is or is not considered *normal* for a population (NAEYC, 2020).

Pedagogy

A type of teaching style or educational approach to ensure content is more meaningful to students by implementing research-based knowledge (NAEYC, 2020).

Standardized test

An assessment that contains specific guidelines for administering, scoring, and analyzing academic end-of-year learning goals (standards) based on predetermined student norms. The results are compared to other district scores to determine if the student population is meeting, exceeding, or falling behind on academic expectations (Frans et al., 2020).

The Start of Inappropriate Practices

Research showed a shift away from developmentally appropriate practices and a movement toward federally funded accountability measures, based on norm-referenced testing systems started inappropriate practices (Rushton & Juola-Rushton, 2008). Based on adequate yearly progress (AYP) scores, the government rated schools and determined whether punitive steps needed to be made (Bauml, 2016; Cramer et al., 2018). The following is a look at where the idea for educational reform originated and what influenced those policies.

Assumptions That Led to Counterproductive Educational Acts

Two systems in society have had little to no change in over a century: religion and the educational system (Wolk, 2011). Instead of changing the educational system to accommodate student needs, policymakers pushed for more rigor. There was an opinion that if teachers and students worked harder, schools would succeed (Wolk, 2011). Each year students and teachers were expected to prove curriculum mastery and school success based on standardized tests. The tests did not follow developmentally appropriate research but instead next-grade level standards (Gullo & Hughes, 2011; Im, 2017; Miller & Almon, 2009; NAEYC, 1995). Early childhood educators value academic rigor and challenging students to reach full potential, but educators also recognize the need for balance in rigor and developmentally appropriate practices.

The *No Child Left Behind Act of 2001* stated that it was “An Act: To close the achievement gap with accountability, flexibility, and choice so that no child is left behind” (2002, p. 115 Stat. 1425). The act was an amended version of the “Title I of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 6301 et seq.)” (2002, p.115 Stat. 1427). Throughout the *No Child Left Behind Act of 2001*, there were broad statements on the implementation of guidelines and distribution of funds pending submitted assessment results. The act briefly

addressed English language learners, stating that students had a strict time limit of three years access to language assistance for standardized tests (2002, p. 115 Stat. 1451). All federally funded and required assessments were to be administered to all students similarly. Even if there were some accommodations available, it was the school district's responsibility to access them, and during the research of this thesis, the directions for how to access those accommodations were challenging to find.

Effects of These Counterproductive Educational Acts on Students

Stephen Rushton and Anne Juola-Rushton (2008) focused on how the pressure to meet standards caused educators to shift away from developmentally appropriate practices and move toward teaching what would be on tests. Edward Miller and Joan Almon (2009) discuss how policies like *the No Child Left Behind Act of 2001* were well-intended but fundamentally flawed because the practices did not come from evidence-based research. Those well-intended ideas should have considered the current demographics of the public-school system. Students enter classrooms with limited English and are required to take tests on iPads or computers with no interpretations available. Cathy Yun, Hanna Melnick, and Marjorie Wechsler (2021) discussed how computer-based or multiple-choice assessments failed to capture children's mastery of content but instead assessed technology exposure or test-taking skills.

In a 1995 position statement by the National Association for the Education of Young Children (NAEYC), there was a discussion on kindergarten expectations pushed to first-grade expectations and widening the achievement gap. NAEYC (1995) stressed the limited view of learning captured on standardized tests and how children were inaccurately labeled as 'high-risk' or 'inadequate' when children were developmentally on track and within the developmental continuum (NAEYC, 1995; Blessing, 2019; Yun et al., 2021). NAEYC (1995) suggested that

schools respond to diverse learners' classrooms by providing curriculum with assessments built into lessons and activities that observe the whole child. The inaccurate labeling of student achievement based on standardized testing causes unnecessary stress on families and educators. A view of how that stress effected educators will be discussed next.

Effects of These Counterproductive Educational Acts on Teachers

Since the adoption of accountability measures based on standardized testing scores, there has been a steady rise in teachers leaving the educational system (Hine et al., 2022; Perryman & Calvert, 2020; Ryan et al., 2017). Research showed the top three reasons for educators leaving the teaching profession were standardized testing pressures, excessive workloads, and lack of work/home balance leading to declined teacher well-being. Even though studies showed that accountability measures did not change how teachers felt about students, accountability measures affected feelings about continuing in the field of education (Perryman & Calvert, 2020). Teachers started the profession believing in the power of quality education. A significant reason educators cited for entering the field was a previous teacher that had inspired a love of teaching and learning (Perryman & Calvert, 2020). The same educators Perryman & Calvert (2020) surveyed and cited a passion for teaching ultimately decided to leave the profession because the pressures led to burnout.

What Is Considered Developmentally Appropriate

The National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) (2003) created a joint position statement on quality early childhood assessments, curriculum, and programs. Valuable information for early childhood educators included: the importance of advocating for developmentally appropriate practices; the need to value growth

over specific data points; the benefit of ongoing professional development for educators of multilingual families; the need for curriculum and assessments that are relevant to all aspects of children's lives (NAEYC & NAECS/SDE, 2003).

Lawrence Schweinhart (2006) researched participatory learning and programs that follow the theories of Jean Piaget. According to Schweinhart (2006), participatory learning is about how children learn versus the content learned. The early childhood education years are when children's most academic, social, and emotional growth takes place, and those skills deepen through participatory learning (Schweinhart, 2006). With participatory learning, it is vital to find a balance between academic rigor and developmentally appropriate practices (NAEYC, 2020; Schweinhart, 2006; Rushton & Juola-Rushton, 2008). The research conducted by Schweinhart (2006) showed resources and training prepared early childhood educators for the vast array of developmental stages in children and ensured proper assessment of growth within early childhood classrooms.

Conclusion

How can urban early childhood educators advocate for accountability in alternative assessment methods? Evidence showed educator knowledge of child development and how children authentically exhibit learning as powerful tools in advocating for alternative assessment methods (Cramer et al., 2018; Frede & Ackerman, 2007; Gullo & Hughes, 2011; Miller & Almon, 2009; NAEYC, 2020; Schweinhart, 2006; Rushton & Juola-Rushton, 2008). Educators understand the need for accountability to ensure quality education for all students. The struggle for educators has been the accountability measures used by policymakers. The current accountability measurements in place do not accurately represent the vast demographics of current classrooms, do not measure student growth on a continuum, and come from a single

assessment; which is not considered developmentally, linguistically, or culturally appropriate (Cramer et al., 2018; Gullo & Hughes, 2011; Miller & Almon, 2009; NAEYC & NAECS/SDE, 2003; Schoenbeck, 2000; Schweinhart, 2006; Wolk, 2011; Yun et al., 2021). The following literature review examined whether standardized testing led to instructional improvement, how to advocate for accountability in alternative assessment measures by clarifying what constitutes developmentally appropriate practice (DAP), the ethical and professional responsibility of educators to advocate for DAP, different ways that teachers adjusted pedagogy to meet standards more appropriately, and research-based options for alternative assessments educators could reference and share with administrators.

Chapter Two: Literature Review

There is tremendous importance in how educators assess students, use assessment information, and help students reach full potential (NAEYC, 2020). To meet the vastly different needs of students, educators regularly assess student understanding and reflect on what supports need to be in place to help each student attain content mastery. Assessing students provides valuable information when used for guiding instruction. The introduction of standardized testing as a form of accountability made assessments less beneficial and more challenging for students and educators (Bauml, 2016; Cramer et al., 2018; Frans et al., 2020; Gullo & Hughes, 2011; Grisham-Brown et al., 2008; Im, 2017; Rushton & Juola-Rushton, 2008). Research has shown that since the introduction of mandated standardized testing by the No Child Left Behind Act of 2001(2002) and the implementation of mandated frameworks in other countries (Mueller Gathercole et al., 2008; Pyle & DeLuca, 2013), there was a shift from assessing to inform educational practices to assessing for mandated accountability (Bauml, 2016; Grisham-Brown et al., 2008; Gullo & Impellizeri, 2021; Rushton & Juola-Rushton, 2008). The pressure to meet mandated requirements while meeting the needs of individual students is a serious task. Educators have shown how creative they can be, how vital early childhood training is for the educational system, shed light on the need to address the stress and pressure standardized testing brings to classrooms, and how to work through the pressures to help future educators (Bauml, 2015; Gullo & Impellizeri, 2021; Im, 2017; Pyle & DeLuca, 2013; Ryan et al., 2017).

This literature review analyzes and synthesizes a combination of qualitative, quantitative, and mixed-method studies along with guidelines from the National Association for the Education of Young Children's [NAEYC] (2020) Position Statement on Developmentally Appropriate Practices (DAP) to examine research on whether standardized tests improve instruction, the

effect of teacher and student views on assessments, what constitutes developmentally appropriate practices, how educators faced accountability challenges, and finally how educators can use this research to advocate for alternative assessment methods that show accountability within urban early childhood classrooms.

Does Standardized Testing Improve Instruction

Considering the rise in standardized testing, Haesung Im (2017) examined whether standardized tests influenced reading scores or if the amount of time spent testing or conducting interventions improved instruction results. The quantitative study by Im (2017) used four-point Likert scale questionnaires, average weighted IRT reading scores, multi-structured equation models, and other variables or covariates to determine whether more standardized testing led to more reading instruction and better test results; or if the type of reading instruction influenced results. The random chance of a student being at a school providing different curriculum added to the experimental value of the quantitative study (Galvan & Galvan, 2017).

Method

The two research questions being investigated by Haesung Im (2017) were: (1) “Is the frequency of standardized testing in kindergarten directly associated with children’s reading achievement at the end of kindergarten, after controlling for student-level and school-level covariates?” (2) “Does the amount and types of reading instruction in kindergarten mediate the relationship between the frequency of standardized testing and children’s reading achievement near the end of kindergarten?” (p.11).

Haesung Im (2017) started with the data use theory by selecting data collected by the *National Center for Education Statistics* (NCES, 2009; cited by Im, 2017) from a national dataset called the “Early Childhood Longitudinal Study- Kindergarten Cohort of 2010-2011

(ECLS-K)", that contained a sample of about 20,000 kindergarten students in the United States during the 2010-2011 school year (Im, 2017). There were multiple steps used to reach the un-weighted results.

There were multiple forms of measurements used in the study. A two-level nested structure, where Level 1 referred to a student variable and Level 2 referred to the school variable. Teachers were given questionnaires and surveys to complete that implemented different point value Likert scales. A three-factor model from a confirmatory factor analysis (CFA) included six whole language items, three phonics items, and three balanced-approach items (Im, 2017). The school-level reading instruction time and frequency were gathered with a 5-point scale.

Control variables were taken into consideration at both the child and school level. School level control variables included: if the school was public or private, how many students received free/reduced lunches, proportions of minority students, location of the school, average teacher degree levels, average years of experience that educators had, and test scores (Im, 2017). The student level control variables included: socioeconomic status (SES), initial reading score, race, gender, age in month, home language, and whether students had an IEP or qualified for special needs services (Im, 2017).

Finally, data was analyzed starting with descriptive statistics and then by using a multilevel structural equation model that used variables as the mediator (Im, 2017). There was an equation for the student level and school level and all paths were adjusted on both levels using the different variables. Im (2017) used remediation to account for the possibility of any error or bias.

Although the information and calculations were hard to follow. The multiple measurements used displayed a thorough investigation and identified efforts made to cut down the chance of error or bias in the findings, by considering all the possible outlying factors.

Results and Discussion

There were five key findings noted: (1) children in kindergartens that administered standardized tests often did not perform better, (2) more frequent testing correlated with more reading instruction time, (3) pedagogical knowledge effects how assessment data is used by educators, (4) only balanced approach reading instruction correlated with frequent assessments and reading success, (5) more standardized tests do not benefit children's learning if educators are not using effective reading instruction (Im, 2017).

Limitations

Haesung Im (2017) noted that the study's limitations were the secondary data analysis and the inability to control what was first measured. Because of this, aspects are missing that could make a difference in the calculations, such as there were no clear definitions of standardized testing provided to or given by educators, there was no detailed information on instructional practices, and teachers' views on classroom practices versus actual practices did not get observed or calculated. The absence of this data in these areas led to subjectivity (Im, 2017).

Another area noted was that in the original study performed back in 2010-2011, the balanced approach was "a novel type of reading instruction" and was chosen by some educators because it aligned more with Common Core (Im, 2017, p.16). That factor could influence the number of educators that used the approach, affecting the sample size and outcomes.

Limitations to the study were hard-to-follow data points and equations. The tables and written analysis could have been more straightforward. Demographics were vague, but it did

mention that those factors were variables added to the calculations. How detailed those demographics were within the actual database used is unknown. Demographic data were drawn randomly from a large pool within the database (Im, 2017).

How Educator Experiences with Mandated Tests Effects Views on Test's Purpose

Educators understand the value of assessing children's knowledge and teachers are a big part of all forms of assessment in the educational system. The mixed-method study by Frans et al., (2020) used the following guiding questions: 1) To what degree do early childhood educators view a norm-referenced tests as an instrument that can serve the purpose of improvement and/ or accountability; 2) Which aspects play a role in the differing experiences that teachers have of standardized (norm-referenced) testing?

Methodologies

Frans et al. (2020) reached out to schools that had completed a research study for them in 2018. The main contact person at the different schools received an email containing a questionnaire asking for the email be forwarded to the special services coordinators, preschool teachers, and kindergarten teachers. Data on participants age, gender, position at the school, years of experience, and email address were requested. Participants were asked to complete an informed consent form before completing the questionnaire. The beginning pool had 97 participants. Researchers used the "Conceptions of Assessment Abridged questionnaire (CoA-III-A; Brown, 2006; cited by Frans et al., 2020, p.92) and told participants to answer based on personal views of assessments designed by "Cito" (Frans et al., 2020, p.92). Once data was collected then 6 final participants were chosen. Participants were selected based on ranks within the Mokken scale scores. The goal was to get a group of participants that had maximum variation to ensure a wide array of viewpoints were collected (Frans et al., 2020). Once the final 6

participants were chosen, the team compared the demographics of the participant's school with the public Netherlands and Dutch primary school's demographics.

Multiple interviews were conducted, recorded, and transcribed. Research team members were referred to as "first author" (Frans et al., 2020, p.94). An undergraduate student transcribed the recordings. Transcriptions were compared to the original recording to check for accuracy and then sent to the participant for *member checking* (Galvan & Galvan, 2017) to see if there was anything they wanted to change. Members could also use the transcriptions to prepare for any upcoming interviews (Frans et al., 2020). Emergent themes from the field interviews were discussed as a team and put together, then coded to create a cluster graph covering the data (Frans et al., 2020).

Results/Researcher Discussion

After analyzing the interviews, research showed that teachers agreed the purpose of the test is accountability and improvement, and either believed it was suitable for that purpose or not. The area where a broad difference of opinion arose was in how participants experienced or coped with testing purposes (Frans et al., 2020). Teachers that viewed the tests as a way for students to gain exposure to standardized testing, were not affected by the scores that children received. Teachers that consistently had students perform poorly on the tests or showed minimal growth, believed poor performance would continue because the teachers felt that standardized tests would never show what students knew, and were a contradiction of the progress observed in the classroom (Frans et al., 2020).

Another aspect that influenced positive or negative emotions went along with how administrators viewed the test results or how administrators supported staff. Positive interactions

with administrators increased the positive mindset towards results and the test's purpose (Frans et al., 2020).

Limitations

The authors of the study stressed that the findings were not meant to evaluate the CoA-III test in any way and would not be sufficient for anyone to use it as such. The low values in the study make it an “approximate ranking of participant” and therefore “limits the utility of these scales for other purposes where more precision is required” (Frans et al., 2020, p. 102). Since a smaller version of the questionnaire was used, the full questionnaire could be used instead to get some other valuable information (Frans et al., 2020).

As far as flaws found during analysis, there were only a few to note. One would be that the demographics of the participants was not shared, nor was there clear discussion about the school's demographics for others to reference. Another critique would be that their entire paper backed up the leading questions, but the conclusion referenced research that had not been discussed anywhere else in the paper and was a strange way to end a strong study.

Link Between Student's Approach to Learning, Teacher Ratings, and Academic Success

According to Gullo & Impellizeri (2021), there has been varying research on the link between a student's behaviors and characteristics on future academic success. This observation led Gullo & Impellizeri (2021) to research how kindergarten teachers' ratings of children's end-of-year math, literacy, and approach to learning (ATL) could predict future elementary success. The focus was the first two years of standardized testing in third and fourth grade, along with which academic or behavior competencies were most predictive of long-term success. A five-year quantitative study examined how effective teachers' observational assessments were at predicting long-term academic success.

Using data from the "Early Childhood Longitudinal Survey- Kindergarten Class of 2010-2011 (ECLS-K: 2011; Tourangeau et al., 2018; as cited in Gullo & Impellizeri, 2021, p. 304), followed by determining dependent variables (end of year 3rd and 4th-grade achievement scores) and independent variables (teacher's ratings from Academic Rating Scale), researchers implemented the "hierarchical multiple regression (HMR)" model, and determine which order predictor variables would enter into equations (Gullo & Impellizeri, 2021, p. 304).

The study sample demographics included: 51% males and 49 % females; 53% Caucasian; 25.3% Hispanic; 13% Black; 8.5% Asian; 4 % two or more races; and 1% Native American/Pacific Islander (p.304). Most students spoke English at home; 18.3% non-English; 1.1 % were multilingual (p. 304). The socioeconomic status of sample participants ranged from 21.3 & below the poverty level to 54% at or above the poverty level (Gullo & Impellizeri, 2021).

Interesting findings in the study were: that socioeconomic status was a more powerful predictor of third and fourth-grade literacy or math performance than gender (Gullo & Impellizeri, 2021). Approach to learning (ATL) was a powerful predictor of math success in 3rd and 4th grade, and math skills like sorting and solving problems were more significant predictors of math success than fraction knowledge. Ultimately, the basic kindergarten learning standards like identifying upper- and lower-case letters and the ability to solve problems (both mathematical and social-emotional problems) could predict long-term success.

According to the Gullo & Impellizeri (2021) research, teachers' appraisals were strong predictors of students' long-term academic success. The research could allow for more in-depth and research-based advocacy of using alternative assessment methods to show accountability in

urban classrooms and encourage trust in kindergarten teachers' ability to use developmentally appropriate practices to show accountability and predict future academic success.

What Constitutes as Developmentally Appropriate Practices (DAP) in Assessing Children

According to the National Association for the Education of Young Children [NAEYC] (2020), implementing developmentally appropriate assessments requires educators to know about child development across all domains (physical, social-emotional, linguistic), knowledge of academic and behavioral standards, an understanding of authentic forms of formal and informal assessments, and the ability to identify a child's progression through various stages of mastery. Educators have a professional, ethical, and moral responsibility to advocate for students when developmentally appropriate practices are not in place (NAEYC, 2011, 2020). A look at how educators have taken the professional and ethical responsibilities to children seriously and faced the accountability challenge with developmentally appropriate practices (DAP) is discussed next.

How Educators Faced the Accountability Challenge with DAP

Much research has observed how accountability and standardization affected upper grades, but very little on how this has trickled down into early childhood curriculum and assessments (Pyle & DeLuca, 2013; Gullo & Impellizeri, 2021). It is essential to understand what early childhood educators are already doing to connect developmentally appropriate practices and accountability measures (Bauml, 2016; Gullo & Hughes, 2011; Im, 2017; Pyle & DeLuca, 2013). The following studies discuss pedagogy, curriculum adaptations, multi-lingual learner assessment advocacy, and linking authentic assessments to accountability measures.

Individual Pedagogical Approaches

The qualitative study done by Pyle and DeLuca (2013) observed three pedagogical approaches: developmental, blended, and assessment for learning. Three kindergarten educators from the same Ontario school district were referred to the researchers by colleagues. Participants' teaching experience ranged from four years to 22 years. Over a four-month period in 2012, researchers used a two-tiered approach with semi-structured interviews that were audio recorded, and classroom observations to see the connection between individual teaching styles and mandated curriculum requirements (Pyle & DeLuca, 2013).

The interviews and observations revealed that even with three different approaches, subjects agreed that having common academic goals and program standards are important guidelines that ensured all content areas were covered regardless of pedagogy style (Pyle & DeLuca, 2013). One observation in the study showed how the participant with 22 years of experience preferred the developmental approach and only used the standard curriculum as a guide for scope and sequence. The veteran educator strongly felt that teachers needed to have knowledge in child development and ensure that curriculum expectations were tied to that knowledge. The types of assessment used were authentic forms, such as observing and documenting behaviors on a developmental checklist (Pyle & DeLuca, 2013).

The more novice participant (4 years' experience) followed the “*assessment as learning*” approach (Pyle & DeLuca, 2013, p. 378). The novice participant did not stray from the given curriculum and mentioned the requirement to submit both daily and long-range plans to the administrator. A notable observation of the novice educator was how the same value was placed on developmentally appropriate practices as expressed by the veteran teacher. The novice followed the students' lead and used feedback strategies to encourage content learning and metacognitive development (Pyle & DeLuca, 2013; see also Gullo & Impellizeri, 2021; NAEYC,

2020). These strategies included an interesting approach to authentic assessment and “*assessment as learning* principles” (p.378) by video recording classmates retelling a story, followed by having the whole class view the recording and providing feedback to fellow students as a form of peer- and self-assessments (Pyle & DeLuca, 2013).

All the teachers, regardless of years of experience, used developmentally appropriate practices while still adhering to the mandated curriculum. Even the novice educator found how to bring a form of authentic learning to the mandated curriculum and gave the class a unique experience while staying firm with guidelines. Pyle & DeLuca (2013) asserted that although the accountability movement creates standardized assessments and curriculum, that does not standardize teacher pedagogy.

Pyle & DeLuca (2013) noted some limitations to the study, including the small sample size, lack of student feedback, and limited ability to show substantial recommendations for one pedagogy over another. Within the current review of research, the lack of teacher or student demographics was noted, along with the discussion section not covering how the classrooms performed, only providing a general statement that teachers met the standards.

Adapting, Augmenting, or Extending Curriculum

Along with the Pyle & DeLuca (2013) study, teachers within Michelle Bauml’s (2016) mixed-method study also adjusted the curriculum while continuing to meet standardized requirements within different classrooms. The added aspect worth noting to Bauml’s (2016) research is the push for current educators to share strategies with future educators faced with the accountability and standardization challenge.

Participants in Bauml’s (2016) study come from the same urban school district that hosts preservice teachers from the researcher’s university, had required quarterly curriculum-based

assessments (CBAs), and worked with the same standardized curriculum materials that included a pacing calendar. Participants ranged in experience levels of 6-25 years, had a large population of English language learners, the student body included many from low socioeconomic backgrounds, and the school had to be ranked as acceptable or above based on mandated academic scores (Bauml, 2016).

Over the 2012-13 school year, participants went through an ongoing process of interviews, observations, and debriefing (Bauml, 2016). Interviews were recorded and transcribed, then entered a QSR NVivo 9 software program that creates descriptor codes to show frequently used words within the interviews. The most frequently used words were *adapted*, *augmented*, and *extended*, and were the focus of the study (Bauml, 2016).

Although all the teachers understood the need to adhere to district requirements, the rigor and pace increase was something that required creative thinking. All participants followed student's lead (Bauml, 2016; see also NAEYC, 2020; Pyle & DeLuca, 2013) and adjusted how material was presented while still staying within the timeline requirements. Some participants *adapted* the curriculum by combining easier lessons together so more time could be spent on more difficult content areas. Some participants *augmented* or added more hands-on activities or replaced some of the curricula with theme-based activities that were more linguistically appropriate. Other participants *extended* curriculum to provide more challenging content for students ready for the next step (Bauml, 2016).

The participating teachers in Bauml's (2016) study illustrated the importance of understanding content requirements, and the need for flexibility within that content to meet the needs of individual students. Although some critics could call this practice 'teaching to the test'; it could also be viewed as differentiation. Either way, Bauml (2016) stressed the importance of

sharing strategies with future educators. By sharing what has and has not worked, preservice teachers could learn how to identify the need for making changes within the curriculum to meet student's needs (Bauml, 2016, p. 91; Pyle & DeLuca, 2013; NAEYC, 2020). Bauml's (2016) research showed educators' mindfulness of developmentally appropriate practices. Participants modified curriculum, understood when and how to use different strategies, exhibited flexibility, and willingness to advocate for students when curriculum or assessments were not adequately meeting children's needs (NAEYC, 2020).

The limitations noted by Bauml (2016), including the small sample size, schedule conflicts that caused one participant to not be observed as often as others, and there was no data collected from administrators. Although it was stated that the schools received a satisfactory rating from the state, data on how classrooms compared to other schools could be beneficial to know.

Advocating for Adjustments in Multi Language Learner Assessments in Wales

To advocate for multilingual learners in Wales, Professor Virginia Mueller Gathercole, Enlli Mon Thomas, and Emma Hughes (2008) set out to achieve a norm-referenced assessment that calculated scores based on the general tested language score combined with a home language placement score based on peers with the same home language, age, and language acquisition exposure. Mueller Gathercole et al., (2008) asserted that multilingual students are the norm in society, but assessment measures do not adequately recognize multilingual students' knowledge because the assessments assume that monolingual students are the norm.

Method

Using a sample size of 611 children (age range 7-11; 317 females and 294 males) from different parts of Wales; North Wales ($N=321$), Mid Wales ($N= 211$), and South Wales($N= 79$

(p.686), and different home languages identifiers: *OWH*= only Welsh spoken, *WEH*= Welsh and English, *OEH*= only English spoken; researchers had students complete questionnaires in order to analyze responses to student responses to 240 words (Mueller Gathercole et al., 2008). From those 240 words, researchers eliminated non-discriminating words and ended up with a list of 111 significant words. Researchers completed student raw scores along with creating hypothetical students that were the same age and had the same language acquisition to check for validity and reliability (Mueller Gathercole et al., 2008). There was reliability in the scores and researchers felt that further research should be done to reach the long-term goal of having the analyzed factors included in mandated testing calculations.

Limitations/Discussion

The limitations were the use of a less known language (Welsh) and comparing it with more well-known language such as English. The validity and reliability regarding other more common languages, would require further investigation and was not completed during the study (Mueller Gathercole et al., 2008). The study was an important step in advocating for assessment modifications for multilanguage learners. The quantitative research could open the door to further research and advocacy on behalf of other languages allowing for more equitable assessment practices within norm-referenced assessments (Mueller Gathercole, 2008).

Linking Authentic Assessments to Accountability Measures

In the research conducted by Grisham-Brown et al. (2008), the most developmentally appropriate form of assessment was authentic assessment, but there was a need for educators to create a clear rubric for high-quality authentic assessments that could be used for accountability measures. With the connection to Head Start and understanding of Head Start's approach to early

childhood learning, the researchers wanted to attempt to show fidelity in educators' ratings of authentic assessments in individual classrooms.

The study required intense training, support, and integration of training within the program and its administrators (Grisham-Brown et al., 2008). The technical assistance program was provided by the Project LINK project and Assessment, Evaluation, and Programming System [AEPS] for Infants and Children (2nd ed.; AEPS; Bricker, 2002; as cited in Grisham-Brown et al., 2008, p. 273). AEPS was considered an authentic and developmentally appropriate assessment due to meeting the following criteria: assessments were performed by familiar adults during daily activities; the assessment involved families and transparency; assessment was linked to programming and intervention efforts. Fourteen educators from 9 different Head Start classrooms were evaluated. Thirteen of the fourteen were female educators, six self-identified their race (7 European American, 7 African American, and 1 Latino). Classrooms consisted of 20 children with and without disabilities as well as including English language learners.

Limitations within the study included the small sample size and results were limited to Head Start and Project LINK guidelines. Another limitation noted was the amount of time needed to use the assessment with fidelity. The educators reported using the tool more accurately three years after the quantitative study was completed. The research of other guidelines using this form of assessment would be worth investigating.

Conclusion

Research findings have demonstrated the ability to show accountability in alternative assessments by matching standards to observation criteria, focusing on strengths-based assessments, and exhibiting next-level thinking within authentic assessments (Frede & Ackerman, 2007; Grisham-Brown et al., 2008; Gullo & Hughes, 2011; Gullo & Impellizeri,

2021; NAEYC, 2020). Educators agreed that assessing students is an important piece of teaching and needs to be used to evaluate what students know, and what ways to scaffold learning to encourage growth (NAEYC, 2020; Bauml, 2016; Pyle & DeLuca, 2013). This review demonstrated how educators faced the challenge of accountability within standardized measures and found multiple ways to advocate for individual learners (Bauml, 2016; Grisham-Brown et al., 2008; Mueller Gathercole, 2008; Pyle & DeLuca, 2013). It is important to understand what early childhood educators are already doing to connect developmentally appropriate practices and accountability measures, and to think about how to face standardized testing without completely changing pedagogy styles that have worked, especially as educators start a journey toward advocating for alternative assessments (Bauml, 2016; Gullo & Hughes, 2011; Im, 2017; Pyle & DeLuca, 2013). Using this combination of qualitative, quantitative, and mixed-method studies along with knowledge of developmentally appropriate practices, educators can advocate for alternative, authentic assessment methods to be used to show accountability, and share more accurate measures of what urban classroom students can do. In the next chapter, different forms of authentic assessments will be discussed along with the guiding principles used by the National Association for the Education of Young Children [NAEYC] (2020). With guidance from NAEYC (2020) and resources on alternative assessments, educators will be able to advocate for accountability measures using alternative assessments in urban early childhood classrooms.

Chapter Three: Insights, Application, and Future Studies

The literature review analyzed and synthesized qualitative, quantitative, and mix-method research on standardized tests and the effect of inappropriate testing measures on students and teachers. In this chapter, insights on how the literature review research can be used to advocate for accountability in alternative assessments will be discussed, along with recommendations on alternative assessment measures, and suggestions on future research needed to advocate for policy changes concerning norm-referenced, standardized testing practices. Educators know that standardized tests cannot capture student learning accurately, because there are too many areas of development that are not accounted for in standardized, norm-referenced tests (Frede & Ackerman, 2007; Im, 2017; NAEYC, 2020; Pyle & DeLuca, 2013; Rushton & Juola-Rushton, 2008; Yun et al., 2021). The following is look at the insights gained that could help educators advocate for accountability in alternative assessment measure.

Insights Gained That Advocate for Accountability via Alternative Assessments

The current standardized, norm-referenced format does not assess children in the way their brains learn (Rushton & Juola-Rushton, 2008). Standardized testing does not account for the heightened sensitivity of early childhood student's physical or emotional needs, the continuum that students learn on, the need to measure student learning in multiple contexts, or how norm-referenced assessments tend to penalize children of different ethnic groups (Bodrova & Leong, 2001; Im, 2017; NAEYC, 2020; Rushton & Juola-Rushton, 2008; Yun et al., 2021). The heightened sensitivity to physical and emotional needs also plays an important role in student's future academic success. There is a strong connection between how student's brains are developing executive functioning skills, growth versus fixed mindsets, and children's individual approach to learning (Center on the Developing Child at Harvard University, 2011; Gullo &

Impellizeri, 2021; McCutchen et al., 2015). Research has shown that student's growth versus fixed mindset or positive approach to learning has significant effects on future academic success (Gullo & Impellizeri, 2021; McCutchen et al., 2015). According to the Center on the Developing Child at Harvard University (2011), another important skillset linked to children's academic success is executive functioning skills such as the ability to focus, follow multi-step directions, control impulses, and problem-solving skills. These are skills that do not come naturally and require positive guidance and practice. These social-emotional skillsets along with academic learning goals are not covered by standardized tests and should be accounted for when assessing students' true ability (Center on the Developing Child at Harvard University, 2011; Bodrova & Leong, 2001; Gullo & Impellizeri, 2021; Im, 2017; McCutchen et al., 2015; NAEYC, 2020; Rushton & Juola-Rushton, 2008; Yun et al., 2021). The following is a deeper look at the research on what assessments need to include.

How to Identify Developmentally Appropriate Assessments

Developmentally appropriate assessments measure learning objectives on a continuum, inform teaching, are done during children's active learning, encourage 'whole child' development that provides greater executive functioning growth for long-term retention, is given in the language students are loved in by familiar adults, and is backed by research (Center on the Developing Child at Harvard University, 2011; Bodrova & Leong, 2001; Frede & Ackerman, 2007; Gullo & Hughes, 2011; Mueller Gathercole et al., 2008; NAEYC 2020; Rushton & Juola-Rushton, 2008; Yun et al., 2021). Authentic forms of assessment provide the above recommendations that are developmentally, culturally, and linguistically appropriate for all students and provide multiple ways to show learning.

Authentic Forms of Assessment

The components of authentic assessments include planned and intentional monitoring, observation, teacher-student interaction, analyzing data collected using multiple assessment tools, providing meaningful feedback to students and families, and linking everything together to inform instructional practice (Bodrova, 2008; LeeKeenan & Ponte, 2018) Types of authentic assessments include observation-based, curriculum embedded performance, and work sampling systems such as student portfolios (Bodrova, 2008; LeeKeenan & Ponte, 2018; Miller & Almon, 2009; NAEYC, 2020; Yun et al., 2021)

Observation-based

Students are observed during daily learning. Teacher's use checklists, running records, anecdotal records, or portfolios with work samples that are electronically captured or kept in binders, and accesses learning across all domains (Blessing, 2019; Bates et al., 2019; LeeKeenan & Ponte, 2018). These forms of assessment can be used daily and capture moments of learning and growth during child's play or interactions with table mates. Having a well-thought-out plan on what educators are looking for throughout domains, helps to make the information gathered more useful and usable (LeeKeenan & Ponte, 2018; Schoenbeck, 2000).

Curriculum Embedded Performance

Assessments that are built into the curriculum. This form of authentic assessment is also observed by the teacher with specific learning goals along a continuum that can be monitored and checked off a checklist, include running records of what is said, and documenting skills with work samples over time (Miller & Almon, 2009; LeeKeenan & Ponte, 2018; Schoenbeck, 2000)

Work Sampling System (WSS)

According to Yun et al., (2021) the work sampling system (WSS) is a portfolio that includes teacher observations. This authentic assessment allows for flexibility with alignment according to individual programs, can be used for children three through third grade, has been linked to above average growth in reading and math, and created a more equitable form of assessment for students of color.

Suggestions for Future Studies and Practices

Based on the research into authentic assessments and developmentally appropriate practices, there is a need for policy makers to understand all domains of early childhood development and how it effects children's learning (Center on the Developing Child at Harvard University, 2011; Bodrova & Leong, 2001; Frede & Ackerman, 2007; Gullo & Hughes, 2011; Mueller Gathercole et al., 2008; NAEYC 2020; Rushton & Juola-Rushton, 2008, Yun et al., 2021). This is especially important with the rise of challenging behaviors and educators learning the teaching profession (Hine et al., 2022; Perryman & Calvert, 2020; Ryan et al., 2017; Schmit, 2020). The following suggestions discuss the importance of social-emotional learning to encourage executive functioning development, the need to limit direct assessments, and the need to limit the weight of standardized tests as a measurement for accountability.

Importance of Social-Emotional Learning

The link between student mindset or how learning challenges are handled by children has a huge effect on future academic success (Gullo & Impellizeri, 2021; McCutchen et al., 2015) Policy makers and decision makers need to prioritize social-emotional learning strategies and growth mindset practices to help students learn how to cope with and work through failure. Children need to know that facing a problem, making mistakes, and trying again are key skills to

long term academic success (Gullo & Impellizeri, 2021; McCutchen et al., 2015). Teachers can encourage growth mindset by giving explicit feedback with a focus on efforts made, and by teaching students that brains are constantly changing and learning with each continued effort (McCutchen et al., 2015).

Limit Direct-Assessments to Two to Three Times a Year

Direct assessments are not a poor form of assessment but need to have a specific purpose and be done only two or three times per year to show growth. These types of assessments need to be performed by a familiar adult that the student knows and feels comfortable with to get the most accurate assessment performance (Yun et al., 2021; Rushton & Juola-Rushton, 2008).

Limiting Standardized Test Weight on Important Educational Decisions

If the government requires documented standardized testing scores, then consideration of the students being tested needs to be accounted for (Frede & Ackerman, 2007; Gullo & Hughes, 2011; Mueller Gathercole et al., 2008; NAEYC 2020; Rushton & Juola-Rushton, 2008, Yun et al., 2021). When children are engaged in high-stakes tests the chemicals in the brain and body resemble that of someone being threatened (Rushton & Juola-Rushton, 2008). Early childhood students in that mindset will not give accurate testing data (Im, 2017; NAEYC, 1995; Rushton & Juola-Rushton, 2008; Blessing, 2019; Yun et al., 2021). If is required to submit standardized test scores, educators must advocate for these scores to not be the only measure observed to determine skills within states, districts, or schools; and ultimately determine funding for children (Cramer et al., NAEYC, 2011, 2020; Rushton & Juola-Rushton, 2008). If the results affect them, then the assessments need to truly represent them (Cramer et al., 2018; Mueller Gathercole et al., 2008).

Conclusion

Standardized testing methods are not the best assessment format for accurate measures of student growth or educator's ability to teach content adequately (Bauml, 2016; Cramer et al., 2018; Frede & Ackerman, 2007; Gullo & Hughes, 2011; NAEYC, 2020; Im, 2017; Yun et al., 2021). Somewhere along the road to accountability, decisions makers confused equality for equity in learning, and pushed aside developmentally, culturally, and linguistically appropriate practices (Bauml, 2016; Cramer et al., 2018; Gullo & Impellizeri, 2021). "School reform that requires anything of "all" children inherently ignores the needs of "each" child" (Cramer et al., 2018, p. 494). Early childhood educators have a responsibility to advocate for students when inappropriate policies are enacted (NAEYC, 2011, 2020). Standardized, norm-referenced testing measures are no way to accurately account for student learning, nor an accurate accountability measure for educator performance (Bauml, 2016; Cramer et al., 2018; Frede & Ackerman, 2007; Gullo & Hughes, 2011; NAEYC, 2020; Im, 2017; Yun et al., 2021). Research based on developmentally, culturally, and linguistically appropriate practices have shown that authentic assessing of children's learning is a valuable tool to document student growth and program accountability (Frede & Ackerman, 2007; Grisham-Brown et al., 2008; Gullo & Hughes, 2011; Gullo & Impellizeri, 2021; NAEYC, 2020). The analysis, synthesis, and reflection on the above qualitative, quantitative, and mixed-methods studies gives educators the information needed to advocate for accountability in alternative measures and bring developmentally appropriate practices to classrooms across the globe.

References

- American Federation of Teachers (AFT). (2015). *Elementary and Secondary Education Act*.
American Federation of Teachers. <https://www.aft.org/position/elementary-and-secondary-education-act>
- Bauml, M. (2016). One size never fits all: Teachers' responses to standardized curriculum materials and implications for early childhood teacher educators. *Journal of Early Childhood Teacher Education*, 37(1), 76–95. (Bauml, 2016)
- Bates, C., Schenck, S., & Hoover, H. (2019). Anecdotal records: Practical strategies for taking meaningful notes. *YC Young Children*, 74(3), 14-19.
- Blessing, A. (2019). Assessment in kindergarten: Meeting children where they are. *YC Young Children*, 74(3), 6-13.
- Brown v. Board of Education, 347 U.S. 483 (1954).
<https://supreme.justia.com/cases/federal/us/347/483/#tab-opinion-1940809>
- Center on the Developing Child at Harvard University. (2011). *Building the brain's "air traffic control" system: How early experiences shape the development of executive function: Working paper No. 11*. <https://developingchild.harvard.edu/>
- Cramer, E., Little, M. E., & McHatton, P. A. (2018). Equity, Equality, and Standardization: Expanding the Conversations. *Education and Urban Society*, 50(5), 483–501.
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). <https://www.congress.gov/bill/114th-congress/senate-bill/1177>
- Frans, N., Post, W. J., Oenema-Mostert, C. E., & Minnaert, A. E. M. G. (2020).
Preschool/Kindergarten teachers' conceptions of standardised testing. *Assessment in Education: Principles, Policy & Practice*, 27(1), 87–108.

- Frede, E. & Ackerman, D.J. (2007). Preschool curriculum decision-making: Dimensions to consider. *National Institute for Early Education Research (NIEER)*, 12.
- Galvan, J. L., & Galvan, M. (2017). *Writing literature reviews: A guide for students of the social and behavioral sciences* (Seventh edition). Routledge, Taylor & Francis Group.
- Grisham-Brown, J., Hallam, R. A., & Pretti-Frontczak, K. (2008). Preparing Head Start personnel to use a curriculum-based assessment: An innovative practice in the “age of accountability.” *Journal of Early Intervention*, 30(4), 271–281.
- Gullo, D. F., & Hughes, K. (2011). Reclaiming Kindergarten: Part I. Questions about Theory and Practice. *Early Childhood Education Journal*, 38(5), 323–328.
- Gullo, D. F., & Impellizeri, W. E. (2021). Kindergarten teachers’ ratings of children’s behavioral and learning competencies: Predictive impact on children’s third and fourth grade achievement trajectories. *Early Childhood Education Journal*, 50(2), 301–314.
- Hine, R., Patrick, P., Berger, E., Diamond, Z., Hammer, M., Morris, Z. A., Fathers, C., & Reupert, A. (2022). From struggling to flourishing and thriving: Optimizing educator wellbeing within the Australian education context. *Teaching and Teacher Education*, 115, 103727.
- Im, H. (2017). Kindergarten standardized testing and reading achievement in the U.S.: Evidence from the early childhood longitudinal study. *Studies in Educational Evaluation*, 55, 9–18.
- McCutchen, K. L., Jones, M. H., Carbonneau, K. J., & Mueller, C. E. (2015). Mindset and standardized testing over time. *Learning and Individual Differences*, 45, 208–213.
- Meisels, S. J., Steele, D., & Quinn, K. (1989). *Testing, tracking, and retaining young children: An analysis of research and social policy* [A commissioned paper for the National Center for Education Statistics]. University of Michigan.

Meisels, S. J. (2006). *Accountability in early childhood: No easy answers* (Occasional Paper No. 6; Herr Research Center and Social Policy Erikson Institute). Erikson Institute.

https://www.isbe.net/Documents/meisels_accountability.pdf

Miller, E., & Almon, J. (2009). Crisis in the kindergarten: Why children need to play in school. *Alliance for Childhood*.

Education Code: Prekindergarten-Grade 12, Minn. Stat. § 120-129C (2022).

<https://www.revisor.mn.gov/statutes/cite/120>

Mueller Gathercole, V. C., Mon Thomas, E., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism*, 11(6), 678–720.

NAEYC (National Association for the Education of Young Children). (1995). *School readiness*. Position statement. Washington, DC: NAEYC.

NAEYC (National Association for the Education of Young Children) & NAECS/SDE (National Association of Early Childhood Specialist in State Departments of Education). (2003). *Joint position statement on early childhood curriculum, assessment, and program evaluation building an effective, accountable system in programs for children birth through age 8*. Washington, DC: NAEYC.

National Association for the Education of Young Children (NAEYC). (2019). Professional standards and competencies for early childhood educators. *NAEYC.org*.

National Association for the Education of Young Children (NAEYC). (2020). *Developmentally appropriate practice: A position statement of the National Association for the Education of Young Children*. <https://naeyc.org/dap>

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, 84(2), 113-130

No Child Left Behind Act of 2001, Public Law No. 107-110, 115 Stat.1425- 2094. (2002).

<http://www.congress.gov/>.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

<https://www.congress.gov/107/plaws/publ110/PLAW-107publ110.pdf>

Perryman, J., & Calvert, G. (2020). What motivates people to teach, and why do they leave?

Accountability, performativity, and teacher retention. *British Journal of Educational Studies*, 68(1), 3–23.

Pyle, A., & DeLuca, C. (2013). Assessment in the kindergarten classroom: An empirical study of teachers' assessment approaches. *Early Childhood Education Journal*, 41(5), 373–380.

Rushton, S., & Juola-Rushton, A. (2008). Classroom learning environment, Brain research and the no child left behind initiative: 6 years later. *Early Childhood Education Journal*, 36(1), 87–92.

Ryan, S. V., Von Der Embse, N. P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, 66, 1–11.

Schoenbeck, B. (2000). *Emergent literacy and early childhood education* (2nd ed.). Concordia University, St. Paul.

Schweinhart, L. J. (2006). The High/Scope approach: Evidence that participatory learning in early childhood contributes to human development. *High/Scope Educational Research Foundation*.

U.S. Department of Education. (2016). *Every Student Succeeds Act: Assessments under Title I, Part A & Title I, Part B: Summary of final regulations.*

<https://www2.ed.gov/policy/elsec/leg/essa/essaassessmentfactsheet1207.pdf>

Wolk, R. (2011). *Wasting minds: Why our education system is failing and what we can do about it.* ASCD, 1703 N. Beauregard St., Alexandria, VA 22311.

Yun, C., Melnick, H., & Wechsler, M. (2021). *High-quality early childhood assessment: Learning from states' use of kindergarten entry assessments.* Learning Policy Institute.