

RESEARCH ARTICLE

Can Coaches Be Good Raters of Teacher-Child Interactions in Early Childhood Settings?

Manuela Jimenez

Arizona State University

Bridget Hamre and Jennifer LoCasale-Crouch

University of Virginia

This paper assesses the potential of coaches to provide objective and valid ratings of teacher-child interactions. The study examines the association between a coach-ratings measure, the Teacher Knowledge and Skills Scale (TKSS), and the Classroom Assessment Scoring System (CLASS), an observational measure that has been found to be a valid assessment of teacher-child interactions. The study also examines the association between one possible source of bias, the coach-teacher relationship, and the coaches' ratings of teacher-child interactions. A sample of 152 early childhood teachers and 12 coaches implementing a coaching intervention participated in this study. Results show a strong correspondence between coaches' and observers' ratings of teacher-child interactions, as well as a tendency for coaches to rate teachers with whom they have higher-quality relationships more favorably. The paper discusses possible ways in which the coaches' ratings can be improved and used to have a more reliable, cost-effective way to assess teacher-child interactions.

Keywords: teacher assessment, teacher-child interactions, coaching, ratings, coaches.

As part of the current efforts to improve early childhood education, policymakers and practitioners are searching for reliable, easy to implement, and cost-effective measures of indicators of quality in early childhood education programs (Howes et al., 2008; Zaslow, Tout, Halle & Forry, 2009; Zaslow, Tout & Halle, 2011). In addition, developers of interventions for early childhood programs are often interested in documenting whether or not their efforts to improve teachers' practice have been successful. A study by Sabol, Soliday Hong, Pianta &

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A060021 to the University of Virginia – funding the National Center for Research on Early Childhood Education (NCRECE). The work of the first author was partially supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090002 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence concerning this manuscript should be addressed to Manuela Jiménez, T. Denny Sanford School of Social and Family Dynamics, Po Box 873701 Tempe, AZ 85287-3701. Email: Manuela.Jimenez@asu.edu

Burchinal (2013) found that observational assessments of teacher-child interactions are the strongest predictor of children's learning in early childhood education settings when compared to popular quality indicators such as staff qualifications and staff-child ratio. Observation is often considered the gold-standard in assessment of teacher-child interactions (Cash, Hamre, Pianta & Myers, 2012; Domitrovich et al., 2009; Pianta, 2006; Pianta & Hamre, 2009; Raver et al., 2008). However, observations can be costly and labor-intensive, posing an obstacle for the assessment of teacher-child interactions (Blanton, Sindelar & Correa, 2006; Howes et al., 2008). This obstacle has generated a need for cost-effective assessments of teacher-child interactions.

Coach ratings of teacher-child interactions may be one alternative to costly observational measures. Coaching can be defined as a professional development model focused on providing teachers with individualized support to improve their teaching practice (Boatright, Galluci, Swanson, Van Lare & Yoon, 2008; Neuberger, 2012). In this form of professional development, coaches regularly observe teachers and provide feedback based on their observations, which puts them in a privileged position to assess teacher-child interactions. The present study assesses coaches' potential to provide valid ratings of teacher-child interactions by examining the association between a coach-ratings measure, the Teacher Knowledge and Skills Scale (TKSS, LoCasale-Crouch & Hamre, 2008a), and a widely used observational measure, the Classroom Assessment Scoring System (CLASS, Pianta, La Paro & Hamre, 2008). The CLASS is an instrument that assesses teacher and children's behaviors related to child development and later achievement (Office of Head Start, 2013), and that has been found valid and reliable for the assessment of teacher-child interactions (Burchinal, Vandergrift, Pianta & Mashburn, 2010; Curby et al., 2009; Dominguez, Vitiello, Maier & Greenfield, 2010; Domitrovich et al., 2009; La Paro et al., 2009; Sabol et al., 2013). This has led the Office of Head Start to include the measure as part of the monitoring process of its programs. The TKSS and the CLASS were both developed based on the Teaching Through Interactions framework (Hamre et al., 2013), which allows the comparison of ratings of the same types of interactions by different raters. The study also looks into coaches' ability to provide objective ratings of teachers' interactions with children by examining the association between one possible source of bias, the coach-teacher relationship, and coaches' ratings of teacher-child interactions.

Assessing Early Childhood Teachers' Interactions with Students

Investments in early childhood education have significantly increased during the last decades (Zaslow, Tout & Martinez-Beck, 2010). Quality Rating and Improvement Systems (QRIS) play an increasingly important role in ensuring the impact of these investments by assessing several indicators of quality of early childhood education programs. These assessments provide important information that can guide program improvement and help determine whether policy investments in specific indicators of quality have positively impacted children's development and learning (Connors-Tadros & Carlson, 2011; Sabol et al., 2013; Zaslow et al., 2009).

QRIS use several quality indicators to assess early childhood program quality (Connors-Tadros & Carlson, 2011; Sabol et al., 2013). However, the degree to which these indicators predict children's learning varies. Sabol and colleagues (2013) examined different quality indicators and identified observed teacher-child interactions as the indicator that most strongly predicted children's learning, among indicators such as staff qualifications, family partnerships and learning environments. This finding makes it especially important to identify reliable, cost-

effective measures of teacher-child interactions. Although observations are considered the most direct and reliable measure for assessing teacher-child interactions (Cash et al., 2012; Domitrovich et al., 2009; Pianta, 2006; Pianta & Hamre, 2009; Raver et al., 2008), the use of observational assessments may be prohibitive for early education programs due to the high costs and logistical challenges it entails (Howes et al., 2008). The use of observational assessments can be equally challenging for researchers and developers of interventions interested in assessing the effectiveness of their programs in improving teachers' interactions with children. This aspect of observation has driven researchers and policymakers alike to search for reliable, easy to implement, cost-effective measures for assessing teacher-child interactions (Howes et al., 2008; Zaslow et al., 2009, 2011).

To address this need, researchers and practitioners assessing K-16 teachers' practice have previously used reports from informants such as principals (Gray, 2010; Jacob & Lefgren, 2008), students (Kyriakides, 2005; Potvin, Hazari, Tai & Sandler, 2009), and even parents (Ostrander, 1996). Although the use of these reports in teacher evaluation systems has decreased due to concerns about reporters' bias (e.g. principals may resort to un-standardized ways to assess their teachers [Bill and Melinda Gates Foundation, n.d.; Weisberg, Sexton, Mulhern & Keeling, 2009]), research has shown that when using reliable assessments, reporters such as principals and students can provide valid assessments of teachers' behavior (Bill and Melinda Gates Foundation, 2012; Doumen, Koomen, Buyse, Wouters & Verschueren, 2012; Harris & Sass, 2009; Li, Hughes, Kwok, & Hsu, 2012).

Despite their potential validity, some of the reports used in K-16 settings may not respond to early childhood education's specific needs. Directors of early childhood education centers often do not have the time to regularly observe teachers in their classrooms (Arend, 2010; Guernsey & Ochshorn, 2011; Riley & Roach, 2006), thus limiting the reliability of director assessments of teacher-child interactions. Furthermore, although student ratings have been used to assess teacher-child interactions in older grades (Bill and Melinda Gates, n.d.; Kyriakides, 2005; Li et al., 2012; Potvin et al., 2009), there are concerns about using this type of assessment with young children. Young children may be confused by the tasks that assessing their interactions with their teacher would involve, they may have difficulties responding to verbal direction, and they may not respond consistently (National Research Council, 2008, p. 202). Given these limitations, it is important to identify additional sources for reliable reports on early childhood teachers' interactions with children.

Coaches as Raters

In the past decades coaching interventions have begun to gain popularity, particularly in early childhood education. A considerable number of interventions (e.g. Domitrovich, et al., 2009; Driscoll & Pianta, 2010; Pianta, Mahsburn, Downer, Hamre & Justice, 2008; Powell, Diamond, Burchinal & Koehler, 2010; Raver et al., 2008), as well as several QRIS (Isner et al., 2011) have chosen to include this professional development approach in their programs. Coaching can be broadly defined as a model focused on providing teachers with individualized support to improve their teaching practice (Boatright et al., 2008; Neuberger, 2012).

Various approaches to coaching are currently being used to support early childhood teachers. In their review of research on researcher-led coaching and coaching within QRIS, Isner and colleagues (2011) describe some of the ways in which these approaches vary: Coaching

models can have a broad focus on improving teachers' practice or they may have a more specific focus on improving teachers' implementation of a given curriculum. They can also vary in the type of activities prescribed, including activities as diverse as needs assessment, modeling of practices and observation of teachers' practice. These activities can be delivered with a frequency that ranges between once a week to less than once a month, and for periods of time that vary between a couple of months and up to two years. Finally, the interactions between coach and teacher can be done either in-person or remotely (through phone or online, using videotaped observations). Despite this variety, most coaching models meet the definition of coaching provided by Head Start's National Center on Quality Teaching and Learning (NCQTL, 2012), which defines coaching as a cyclical approach that involves planning, observations of teachers' practice in the classroom and reflection and feedback about the practice.

Since most coaching models include repeated observation of teachers' practice during extended periods of time (Domitrovich et al., 2009; Hsieh, Hemmeter, McCollum & Ostrosky, 2009; Isner et al., 2011; Pianta et al., 2008; Raver et al., 2008), coaches implementing these models are in a unique position to assess teachers' interactions with children. Their repeated observations of teachers' practice provide them with a large amount of information about their interactions with children. In this case, the task of coaching itself can increase coaches' ability to reliably assess teachers' interactions with children (Bill and Melinda Gates Foundation, 2012), making coaches' ratings a possible cost-effective assessment of teacher-student interactions.

Relationship Bias

Even though coaches' tasks can set them up to provide cost-effective ratings of teachers' interactions with students, coaches, just like other reporters, are susceptible to bias. One type of bias that has been identified in general performance assessment situations is the dyad-specific, or relationship bias (Hoyt, 2000) which refers to bias attributable to the raters' perception of specific ratees. For instance, a coach may rate higher those teachers who show more commitment to the coaching process, regardless of their observed level of teacher-child interactions. Relationship bias may be of special concern with coaches' ratings because the constant interaction that is fundamental to a successful coaching process may promote closer relationships between coaches and teachers. These relationships may, in turn, bias the coaches' ratings of a given teacher with coaches rating higher the specific teachers with whom they have higher-quality relationships. A similar concern has been brought up by critics of principal ratings to assess teachers, who mention relationship bias as one of the elements that can affect principals' evaluations, with principals rating higher those teachers with whom they have better relationships (Gray, 2010; Harris & Sass, 2009; Jacob & Lefgren, 2008; Ostrander, 1996).

Although there is not much research specifically on relationship bias in principals' ratings, research from other fields shows that positive relationships between supervisors and subordinates can influence supervisors' ratings of the subordinates' performance. This research has found that subordinates with higher quality relationships with their supervisors obtain higher performance ratings than subordinates with lower quality relationships, after controlling for their objective performance (Breuer, Nieken & Sliwka, 2011; Duarte, Goodson & Klich, 1994; Ferris, Munyon, Baski & Buckley, 2008; Lefkowitz, 2000).

A high-quality rater-ratee relationship could also bias ratings by leading raters to provide more accurate ratings (e.g. closer to the observers' ratings) for those ratees with whom they have

higher-quality relationships. Research on performance assessment has found that raters were more invested in the observation process when they were observing employees with whom they have higher-quality relationships (Antonioni & Park, 2001), which can lead to raters providing more accurate ratings for these employees. In coaching, coaches may pay more attention during observations of teachers with whom they have higher quality relationships which could lead to more accurate ratings for this subgroup of teachers.

The Present Study: MyTeachingPartner and the Teaching Through Interactions Framework

The purpose of this study is to assess the extent to which coaches using a specific coaching approach can provide objective ratings of teacher-child interactions that correspond with observers' ratings using a validated observational instrument. Two specific questions were addressed: 1) To what extent are the coaches' ratings associated with the observational ratings of teachers' interactions with students made by trained observers?; and 2) To what extent are the coaches' ratings associated with the quality of the coach-teacher relationship, both independently or in association with the observers' ratings? Knowing if bias affects coaches' ratings would provide an initial assessment of their objectivity, and if needed, would also allow us to develop strategies to reduce it and help coaches improve their ability to provide objective ratings of teacher-child interactions. Based on previous research regarding the role of relationship bias in raters' ratings, we hypothesize that coaches will provide higher and more accurate ratings of teacher-student interactions for teachers with whom they report higher-quality relationships.

The coaching model used by coaches in this study is MyTeachingPartner (MTP; Pianta, Mashburn, et al., 2008). Similar to other coaching models (NCQTL, 2012) this model takes a cyclical approach that involves planning, observations of teachers' practice in the classroom and reflection and feedback about the practice (Pianta, Mashburn, et al., 2008). MTP has been found to have positive impact on teachers' practice and children' learning and development in evaluations of implementations led both by the developers (Downer et al., 2011; 2013; Mashburn, Downer, Hamre, Justice & Pianta, 2010; Pianta, Mashburn, et al., 2008) and by states' agencies (Early et al., 2014).

MTP consists of a web-mediated coaching process in which every two weeks teachers mail their coaches a 30-minute long videotape of their practice in the classroom. After observing the whole tape, coaches provide teachers with written prompts focused on improving teachers' observation of their own practice, and supporting the analysis of specific teacher-student interactions seen in the video. Teachers' responses to these prompts, as well as any other concerns that the teacher may have, are discussed during a following conference (via phone), where coach and teacher also come up with an action plan for the next video recording.

MTP is grounded in the Teaching Through Interactions framework (TTI; Hamre et al., 2013), a research-based model of effective teacher-child interactions. The TTI framework focuses on three domains of teacher-child interactions: (a) emotional support, which includes how teachers promote social and emotional functioning in the classroom; (b) classroom organization, which includes "processes related to the organization and management of children's behavior, time, and attention in the classroom" (Pianta, La Paro & Hamre, 2008, p. 3); and (c) instructional support, which encompasses teachers' efforts to promote learning in their classroom. In spite of the theoretical differentiation between these three domains, recent studies

have found that an overall factor including elements from all domains predicts children's outcomes across developmental domains (Hamre, Hatfield, Pianta & Jamil, 2014). The coaches' rating measure (TKSS) and the observational measure (CLASS) used in this study are also based on the Teaching Through Interactions framework.

METHOD

Participants and Settings

The data for this study were collected as part of a larger intervention aimed at improving teacher-child interactions for early childhood teachers. The intervention was implemented in nine sites across the United States. The present study includes data on 152 teachers who participated in the coaching condition of the intervention, and their 12 coaches. Demographic statistics are presented in Table 1. Each coach worked with a group of 10 teachers in average (range from 5 to 14) in the span of a school year. The number of coaching cycles a teacher completed varied, with teachers completing an average of 10 cycles (range from 1 to 21).

TABLE 1
Demographic Statistics for Participant Teachers and Coaches

	Teachers (n = 152)	Coaches (n = 12)
Gender (female)	91.4%	100%
Race/Ethnicity		
African American	43.2%	16.67%
Caucasian	35.7%	83.33%
Hispanic	14.6%	-
Asian	1.9%	-
Other	3.1%	-
Highest education		
AA or less	39.1%	-
Bachelor's	37.2%	-
Master's	23.7%	75%
Education specialist	-	16.67%
Doctoral degree	-	.08%
Teaching experience		
Mean (SD)	14.56 (9.56)	13.67 (9.71)
Range	0 - 43	1.5 - 32
Teacher at		
Head Start	50.6%	
Public school	36.4%	

Measures

Observational measure of teacher-child interactions. The Classroom Assessment Scoring System, CLASS (Pianta, La Paro & Hamre, 2008), is an observational instrument scored from 1 (*low*) to 7 (*high*) that assesses teacher-child interactions (Hamre et al., 2013). Previous research using the CLASS has shown the predictive validity of this measure in relation to children's academic, language and social skills (Curby et al., 2009; Mashburn et al., 2008; Sabol et al., 2013).

For the purpose of this study, a team of observers trained on the CLASS coded the videos submitted by teachers for the coaching process. Training consisted of a presentation of short videos that illustrated the CLASS dimensions, followed by a practice coding of five master-coded videos. In order to be able to code, observers had to pass a reliability test in which they needed to score within one point of the master code in 80% of the scores for five videos. While observers were coding they attended weekly meetings to avoid drift on their codes due to rater bias. Observers coded the first 30 minutes of videos submitted by teachers. All segments were double coded and inter-rater reliability was conducted across all footage, with intra-class correlations (ICCs) calculated at the video level ranging from .42-.51. Although these ICCs are not ideal, the use of multiple scores within a certain time frame can improve their reliability. This study used the overall CLASS score at the end of the intervention, which averaged the scores for videos sent by teachers during the last four months of the intervention year (between March and June). Previous studies using the CLASS have found similar inter-rater reliability statistics (Mashburn et al., 2008). At the same time, the data analyzed in the present study has been found to be sensitive enough to identify impacts of MTP on teachers' interactions with students (Downer et al., 2013). Internal consistency was calculated using the scores for the three CLASS domains, resulting in an alpha of .83, showing good internal consistency of the measure.

Coach ratings of teacher-child interactions. At the end of the intervention coaches were asked to complete the TKSS (LoCasale-Crouch & Hamre, 2008a) for each of the teachers with whom they worked. This 22-item scale measures coaches' perceptions of the quality with which a teacher is interacting with children in the classroom. The measure includes items from each one of the three domains of the TTI framework, such as "Teacher provides kids with comfort and assurance" and "Teacher gives kids hints when they can't figure the response out". For this scale coaches were asked to select the response that reflected how much they had seen the teacher engaging in the specified behavior on a 5-point Likert scale in which 1 was *Never* and 5 was *Very Frequently*. This measure showed high internal consistency, with an alpha of .967.

For this study, coaches were asked to rate their teachers' interactions with students based on their observation of the same videos used by the observers for their CLASS ratings. However, the TKSS asked coaches to rate teachers with students based on their recollection at the end of the coaching process of what happened in the videos, while the observers provided their ratings immediately after watching each individual video.

Coach-teacher relationship. The Teacher-Coach Relationship Scale (TCRS, LoCasale-Crouch & Hamre, 2008b) was used to assess coaches' perceptions of their relationship with the teacher. This 7-item scale was adapted from a previous version used in research with preservice teachers. The measure is set on a response scale from 1, *Strongly Disagree*, to 5,

Strongly Agree. Examples of the items include “I have a good relationship with the teacher,” “I am comfortable sharing ideas with the teacher,” and “Interactions with the teacher leave me annoyed and frustrated” (reverse item). The internal consistency of the TCRS in this study was high ($\alpha = .92$)

Data analysis

A hierarchical linear regression was conducted to examine the degree to which the observational ratings of teacher-child interactions (measured by the CLASS) and the quality of the coach-teacher relationship (measured by the TCRS) predicted the coaches’ ratings of these interactions (measured by the TKSS). This regression controlled for the coaches’ years of education and of teaching experience, and the number of coaching cycles in which the teacher participated. To account for the fact that teachers are nested in coaches, the analyses were run in HLM7 Student with a two-level model in which teachers were nested within coaches. The first step of the regression added the observational ratings of teacher-child interactions to the model and the second step added the quality of the coach-teacher relationship. To test the hypothesis that coaches provide more accurate ratings of teachers with whom they have a higher-quality relationship the third step added an interactional term between the observers’ ratings and the coach-teacher relationship.

RESULTS

The goal of this study was to assess coaches’ ability to provide objective ratings of teacher-child interactions that correspond with a previously validated observational measure. The associations between the coaches’ ratings of teacher-child interactions and several coach, teacher and dyad characteristics were assessed in order to identify variables that should be included in the model as covariates to improve the reliability of the findings. Descriptive statistics and bivariate correlations are presented in table 2. Only coaches’ characteristics were found to be significantly associated with coaches’ ratings of teacher-child interactions, with coaches with more years of education providing higher ratings of their teachers’ interactions with students. Consequently only the coaches’ characteristics were included as covariates in subsequent models. Because the number of each teacher’s videos that the coach had access to depended on the number of coaching cycles in which the teacher participated, this variable was also included as a covariate in the model. Correlations also show that coaches’ ratings have a moderate positive correlation with both the observers’ ratings and the coach-teacher relationship. Teachers that received higher ratings on their interactions with students from their coaches also received higher ratings on their interactions based on observations of their practice. Coaches also provided higher ratings for the quality of their relationship with these teachers.

TABLE 2
Descriptive Statistics and Correlations.

	Mean (SD)	Correlations									
		1	2	3	4	5	6	7	8	9	10
1. Coaches' years of education	18.47 (.75)	-	-.252**	-.016	.007	-.250**	.080	.052	-.002	.276**	.303**
2. Coaches' years of teaching experience	13.67 (9.71)		-	-.120	.092	.069	-.299***	-.054	-.026	.079	.053
3. Number of coaching cycles	10.86 (3.58)			-	-.036	.326***	-.014	.231**	.384**	.526**	.389**
4. Teachers' years of teaching experience	14.56 (9.56)				-	-.104	.464***	-.044	.078	.106	.045
5. Teachers' years of education	15.84 (1.69)					-	-.045	.100	.250***	.157	.139
6. Difference in age (teacher – coach)	-5.23 (14.71)						-	-.054	-.043	-.030	-.154
7. Difference in ethnicity (1 = match)	.32 (.47)							-	-.023	.127	.122
8. Observational ratings of teacher-child interactions	4.48 (.53)								-	.438**	.570**
9. Coach-teacher relationship	4.27 (.83)									-	.552**
10. Coach ratings of teacher-child interactions	3.614 (.76)										-

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$

Associations between coaches' and observers' ratings

The first research question examined the extent to which coaches' ratings of teacher-child interactions are associated with observational ratings in a validated instrument. Analyses found a strong correspondence between observers' and coaches' ratings of teacher-child interactions, even after adding the quality of the coach-teacher relationship to the model (see Model 3 in table 2, $\beta = .655$, $p < .001$). This means that coaches and trained observers provided similar ratings of teachers' interactions with children, even after taking into account the association between the coaches' relationship with teachers and their ratings.

TABLE 3
Prediction of Coach Rating Scores of Teacher-Child Interactions (TKSS) by Observed Ratings (CLASS Scores) and Quality of Coach-Teacher Relationship (TCRS), and the Interaction between CLASS and TCRS.

	Model 1 ^a	Model 2 ^b	Model 3 ^c	Model 4 ^d
	β (SE)	β (SE)	β (SE)	β (SE)
Intercept	-4.253 (1.977)	-3.530 (2.235)	-1.750 (2.419)	-1.828 (2.451)
Level 1 (teacher)				
Number of coaching cycles	.085 (.015) ^{***}	.043 (.014) ^{**}	.019 (.016)	.019 (.016)
Observational ratings		.736 (.090) ^{***}	.655 (.092) ^{***}	.670 (.092) ^{***}
Coach-teacher relationship			.209 (.072) ^{**}	.242 (.075) ^{**}
Observational ratings x coach-teacher relationship				.154 (.106)
Level 2 (coach)				
Coaches' years of education	.366 (.058) ^{**}	.353 (.120) [*]	.273 (.129)	.276 (.131)
Coaches' years of teaching experience	.015 (.007)	.013 (.009)	.009 (.009)	.009 (.009)
Deviance	320.968	271.625	264.978	267.368

^a Unconditional model, includes only covariates. ^b Adds covariates and observational ratings to the unconditional model. ^c Adds the coach-teacher relationship to model 1. ^d Adds the interaction between observational ratings and coach-teacher relationship to model 2.

* $p \leq .01$. ** $p \leq .001$

Role of coach-teacher relationship

The extent to which coaches' ratings are associated with the quality of the coach-teacher relationship was assessed in models 3 and 4. Model 3 includes the quality of the coach-teacher

relationship and observers' ratings of the teacher-child interactions, while Model 4 adds an interactional effect between the quality of the coach-teacher relationship and the observers' ratings.

The results of Model 3 show that there is a significant relation between the quality of the coach-teacher relationship and coaches' ratings, confirming our hypothesis of a tendency of coaches to report higher levels of teacher-child interactions when they have a higher-quality relationship with a teacher. However, the results of Model 4 show that there is no significant association between the interaction of the quality of the coach-teacher relationship and the observers' ratings, and the coaches' ratings. This means that coaches tend to give higher ratings to teachers with whom they have a higher-quality relationship, regardless of the level of effective teacher-child interactions identified by an objective observer. These findings provide evidence to reject the hypothesis that coaches provide more *accurate* ratings to teachers with whom they have a higher quality relationship.

DISCUSSION

The goal of this study was to establish whether coaches' ratings of teacher-child interactions could be a cost-effective alternative to observational instruments, by providing objective ratings that are associated with trained observers' ratings of these interactions using a validated observational instrument. The study provides evidence that coaches using the TKSS can provide relatively accurate ratings of teacher-child interactions when compared with observers' ratings. Coaches' ratings on the TKSS showed a strong correspondence with observers' ratings, and the degree of correspondence (i.e. the accuracy of coaches' ratings) was not found to vary depending on the quality of the coach-teacher relationship. This correspondence, however, was not perfect. Results show that coaches tend to rate teachers' interactions with children higher when they have a higher-quality relationship with them. In summary, coaches' ratings corresponded well with observers' ratings and this association was the same across teachers, regardless of the quality of the coach-teacher relationship. This finding provides support to the use of coaches' ratings to assess teacher-child interactions.

The use of coaches as reporters of teacher-child interactions has several advantages. In most coaching models, classroom observation is already one of the coaches' main tasks, and in these observations the coaches are focused on the teacher's practice. Since coaches are already collecting information about effective teacher-child interactions, a coach-reported measure such as the one used in this study would not be a demanding extra task on the coaches. This, added to the short length of the instrument used, makes the TKSS a time-effective measure for coaches to assess their teachers' interactions with children.

The use of a standardized measure like the TKSS signifies an improvement from other methods historically used for assessing K-12 teachers' performance, such as parental feedback and principals' informal walk-throughs (Davidson-Taylor, 2002; Jacob & Lefgren, 2008; Skretta, 2007). Thus, even though the TKSS was developed to capture coaches' ratings of teacher-child interactions, this measure could also be used to more reliably capture other reporters' ratings, such as program directors. By providing a standardized measure, the TKSS could allow directors to focus on the same specific interactions in every classroom, improving the reliability of the assessment process. This study, however, only presents evidence of the coaches' ability to provide objective, valid scores on the TKSS. Further research would be needed to assess if this is

a measure that program directors can use given their constraints to regularly observe classrooms, and whether or not program directors can also provide objective ratings using this measure.

Role of the coach-teacher relationship

In spite of the evidence provided in this study regarding the correspondence between coaches' and trained observers' ratings of teacher-child interactions, coaches' tendency to provide higher ratings to those teachers with whom they have higher-quality relationships could generate doubts about the appropriateness of the use of their ratings. Although further research is required to understand this bias in coaches, research in similar fields could provide an initial approximation to the issue. Previous studies have highlighted the mediating role of the rater's affect in the association between the relationship and the raters' ratings (Judge & Ferris, 1993). Based on this research we could hypothesize that coaches' affect influences the information that they attend to, how they interpret it, how they select it to make judgments and how they recall it (Forgas & George, 2001). This influence could lead coaches to unintentionally interpret their observations in a way that matches their initial perception of the teacher (Duarte et al., 1994). In a way, this bias then may be a result of unavailability of part of the information the coach would need to provide an objective rating of the teacher's interactions with children.

If this was the case, one way to reduce the influence that affect has on coaches' information recall would be to train and support them to use all the relevant information to make the ratings. This support could involve a process where coaches are asked to take detailed notes of their teacher observations using the required framework (in this case the TTI). These detailed notes would provide the coach with a written record of what happened in the classroom, decreasing the need for the coach to recall these interactions and providing a more accurate account of the observed teacher-child interactions. This would allow coaches to base their ratings in specific interactions seen in the observation instead of their general impression after the observation. Similar trainings focused on improving the observational process of raters in other fields have found that such trainings effectively increase the reliability of the reporters' ratings (Kline & Sulsky, 2009; Noonan & Sulsky, 2001; Roch & O'Sullivan, 2003). At the same time, the notes would provide more reliable information for the overall coaching process, which could also help improve its effectiveness.

Another way to increase the precision of the TKSS is to use it in combination with reports from other sources that could complement each other and provide a more complete picture of what happens in the classroom. Previous research has found that multisource assessments are more reliable than single reporter assessments (Bill and Melinda Gates Foundation, 2012; Li et al., 2012). Some QRIS already collect other ratings of teacher-child interactions such as teacher self-report ratings (Howes et al., 2008). These reports, along with ratings from teacher aides or from other teachers that observe the assessed teacher, could be used to complement the coaches' ratings. Since these ratings are already being collected as part of the QRIS they would not be an extra burden on the teachers, maintaining the cost-effectiveness of the measure. The present study provides evidence of how a measure of coaches' ratings of teacher-child interactions could be a valuable addition to a multisource assessment. However, further research would be needed to assess the reliability and validity of other reporters' ratings in early childhood education, as well as their combined value in a multisource assessment.

This study interpreted the significant association between the coach-teacher relationship and the coaches' ratings as a bias in the coaches' ratings due to the quality of the relationship. However, due to the study's design this finding can also be interpreted as coaches engaging in higher-quality relationships with teachers that show higher levels of teacher-child interactions. Higher quality of teacher-child interactions may facilitate the coaching process and increase the opportunities for coaches to provide positive feedback and reinforcement, creating a positive climate in the coaching process that is more conducive to higher-quality coach-teacher relationships. Further research is needed to better understand how a high-quality coach-teacher relationship develops in order to improve our understanding of its association with coaches' ratings.

Practical implications

The present study provides evidence supporting the use of the TKSS, a coach-reported measure of teacher-child interactions. However, this study examined the ratings provided by coaches participating in one specific coaching approach, a researcher-led implementation of a web-mediated coaching intervention. This is only one of several different approaches to coaching that are currently being implemented in early childhood education programs, and some of MTP's differentiating characteristics may limit the generalizability of these findings to other approaches to coaching. Researchers and practitioners should consider these differences if they are interested in using this measure. For instance, in their review of researcher-led coaching and coaching within QRIS contexts Isner et al. (2011) found that researcher-led coaching tends to have more frequent coaching meetings and last longer than coaching within QRIS. These differences may allow coaches in researcher-led implementations of coaching more opportunities to observe and have a clear idea of the interactions between the teachers they are working with and their children, improving their ability to provide reliable ratings.

The generalizability of these findings can also be affected by the training received by coaches in this study. In MTP coaches received training on the TTI framework, which is the basis for MTP, the CLASS and the TKSS. This type of coaches' training, however, is an exception from what coaching interventions within QRIS settings usually do. In most cases, coaches' expertise is assumed to be a prerequisite for the job, and training is not provided. Coaches are usually left to figure out on their own how to implement the coaching intervention (Galluci et al., 2010). Research has found that training can increase raters' accuracy (Cash et al., 2012; Hoyt & Kerns, 1999), so it is possible that coaches not trained in the specific coaching's framework (in this case the TTI) would not be able to provide reports that correspond as highly with observational ratings made based on the framework.

These two issues result from the differences between researcher-led coaching and coaching within the context of QRIS. It is important to note, however, that the landscape of coaching within QRIS is changing, with state agencies starting to use coaching approaches developed by researchers with positive results (Early et al., 2014). In this case the differences between researcher-led implementations and coaching within QRIS may be minimized and findings from this study could be generalizable to the QRIS context.

Finally, although the use of technology in coaching is increasing (Isner et al., 2011), there are still few coaching models that observe teachers' practice via videotapes, and it is possible that the findings from this study would not be generalizable to in-person coaches. The

use of videotapes for ratings of teacher-child interactions may facilitate the rating process by allowing raters to watch an interaction as often as they feel necessary, while raters doing live observations have to be able to focus on the interactions when they are in an environment full of distractions. These distractions could burden coaches' attention and limit the objectivity of their ratings. Although the present study provides evidence that coaches and observers basing their assessments on video observations provide ratings that correspond well, future studies should assess whether this correspondence translates to assessments based on in-person observations.

Because of the importance of the coach-teacher relationship, researchers and practitioners interested in using ratings like the TKSS should also consider the effects on the coaching process of coaches rating the teachers they work with. Introducing this assessment as part of the coaches' task may change the coach-teacher relationship from a helping one to an evaluative one, which may endanger the establishment of a high-quality relationship. To ensure that the assessment task doesn't intervene with the coaching process coaches' ratings should be framed as part of a process to help teachers improve their practice in the classroom and to identify program- or center-wide issues that should be targeted in future professional development efforts.

Limitations

This study has several limitations. Participating coaches assessed teachers' interactions with children using information from several 30-minute long videotapes that teachers chose to send to their coach for feedback. Because teachers select the videos, they can contain bias introduced by teachers' choice of activities or by their desire to appear competent to the coach. At the same time, it is possible that the limited time for observation did not provide enough information for the coach to make a thorough assessment of teachers' interactions with children. However, MTP coaches base the feedback and training provided to teachers in the information gathered from these videos. Given MTP's previous evidence of positive impacts on teachers' practice (Downer et al., 2013; Pianta, Mashburn et al., 2008), the information provided in these videos seems to be sufficient to provide useful feedback and training. This evidence, along with this study's findings of correspondence between coaches' and observers' ratings of teacher-child interactions, supports the idea that these videos provided coaches with enough information for reliable ratings of teacher-child interactions. In this case, the findings from this study could be considered conservative, since the correspondence between coaches' and observers' ratings for coaches that have longer opportunities to observe teachers could be higher than the correspondence identified in this study.

A second limitation in this study concerns the timing of measurement. While observational data were a composite of the last four months of observers' ratings of teacher-child interactions, the coaches' ratings were only collected once at the end of the year. Coaches were asked to base their ratings on the last observations they made of the teachers' practice, which typically varied from March through June. Better alignment between the timing of assessments may improve the level of correspondence between coaches' and observers' ratings.

Summary

The present study provides initial evidence of coaches' ability to provide ratings of teachers' interactions with children that correspond with observers' ratings in a validated observational instrument when basing these assessments on video observations of teachers' practice in the classroom. However, the results also show that coaches rate teachers' interactions with children higher when they have a higher-quality relationship with them. Despite this limitation, coaches' ratings could be a cost-effective option for researchers and practitioners interested in assessing teacher-child interactions. To improve the reliability of coaches' ratings researchers and practitioners could provide coaches with tools to help them use all the relevant information to make their ratings, or they could use coaches' ratings along with ratings of other reporters of the teachers' practice to increase the reliability of the measure. Further research is needed to examine the validity of the suggested multisource assessment in relation to both other validated assessments of teacher-child interactions (e.g. classroom observations), and children outcomes, as well as the effectiveness of coaches' training to increase the reliability of their assessments. Research is also needed to assess the generalizability of these findings to ratings made by coaches within the QRIS context or based on in-person observations.

REFERENCES

- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, 27, 479-495. doi:10.1177/014920630102700405
- Arend, L. E. (2010). *The current state of early childhood education programs: How early childhood center directors manage their human resources*. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses (3411817).
- Bill and Melinda Gates Foundation (no date). *Learning about Teaching. Initial Findings from the Measures of Effective Teaching Project*. Retrieved from: <http://www.gatesfoundation.org/college-ready-education/Documents/preliminary-findings-research-paper.pdf>
- Bill and Melinda Gates Foundation (2012). *Gathering Feedback for Teachers. Combining High-Quality Observations with Student Surveys and Achievement Gains*.
- Blanton, L. P., Sindelar, P. T., & Correa, V. I. (2006). Models and measures of beginning teacher quality. *Journal of Special Education*, 40 2, 115-127
- Boatright, B., Galluci, C., Swanson, J., Van Lare, M., & Yoon, I. (2008) Coaching for instructional improvement: Themes in research and practice. *Washington State Kappan*, 2, 3-5.
- Breuer, K., Nieken, P., & Sliwka, D. (2011). Social ties and subjective performance evaluations: an empirical investigation. *Review of Managerial Science*. Retrieved from <http://www.springerlink.com/content/rw36318525383m8g/>
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25, 166-176. doi:10.1016/j.ecresq.2009.10.004
- Cash, A. H., Hamre, B. K., Pianta, R. C., & Myers, S. S. (2012). Rater calibration when observational assessment occur at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly*, 27, 529-542. doi:10.1016/j.ecresq.2011.12.006
- Connors-Tadros, L., & Carlson, B. C. (2011). Integrating Quality Rating Systems and professional development systems in early childhood. In C. Howes & R. C. Pianta (Eds.) *Foundations for teaching excellence: Connecting early childhood quality rating, professional development, and competency systems in states* (25-46). Baltimore, MD: Paul H. Brookes Publishing
- Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., . . . Barbarin, O. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education & Development*, 20, 346-372. doi:10.1080/10409280802581284

- Davidson-Taylor, C. (2002). Is instruction working? Students have the answer. *Principal Leadership*, 3, 30-35. Retrieved from <http://www.nassp.org/tabid/2043/default.aspx>
- Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A. & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI Program. *American Educational Research Journal* 46 2, 567-597. doi:10.3102/0002831208328089
- Dominguez, X., Vitiello, V. E., Maier, M. F., & Greenfield, D. B. (2010). A longitudinal examination of young children's learning behavior: Child-level and classroom-level predictors of change throughout the preschool year. *School Psychology Review*, 39, 29-47.
- Doumen, S., Koomen, H. M. Y., Buyse, E., Wouters, S. & Verschueren, K. (2012). Teacher and observer views on student-teacher relationships: convergence across kindergarten and relations with student engagement. *Journal of School Psychology*, 50, 61-76. doi:10.1016/j.jsp.2011.08.004
- Downer, J. T., Pianta, R. C., Burchinal, M., Field, S., Hamre, B. K., LoCasale-Crouch, J., ... Scott-Little, C. (2013). *Coaching and coursework focused on teacher-child interactions during language/literacy instruction: effects on teacher outcomes and children's classroom engagement*. Manuscript submitted for publication
- Downer, J. T., Pianta, R. C., Fan, X., Hamre, B. K., Mashburn, A., & Justice, L. (2011). *Effects of web-mediated teacher professional development on the language and literacy skills of children enrolled in prekindergarten programs*. *NHSA Dialog*, 14, 189-212.
- Driscoll, K. C. & Pianta, R. C. (2010). Banking time in Head Start: Early efficacy of an intervention designed to promote supportive teacher-child relationships. *Early Education and Development*, 21, 38-64. doi:10.1080/10409280802657449
- Duarte, N. T., Goodson, J. R., & Klich, N. R. (1994). Effects of dyadic quality and duration on performance appraisal. *Academy of Management Journal*, 37, 499-521. doi:10.2307/256698
- Early, D. M., Maxwell, K. L., Skinner, D., Kraus, S., Hume, K., & Pan, Y. (2014). *Georgia's Pre-K professional development evaluation: Final report*. Chapel Hill, NC: University of North Carolina at Chapel Hill.
- Ferris, G. R., Munyon, T.P., Baski, K. & Buckley, M. R. (2008). The performance evaluation context: Social, emotional, cognitive, political, and relationship components. *Human Resource Management Review*, 18, 146-163. doi:10.1016/j.hrmr.2008.07.006
- Forgas, J. P., & Gorge, J. M. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational Behavior and Human Decision Processes*, 86, 3-34. doi:10.1006/obhd.2001.2971
- Galluci, C., Van Lare, M. D., Yoon, I. H., & Boatright, B. (2010). Instructional coaching: Building theory about the role and organizational support for professional learning. *American Education Research Journal*, 47, 919-963. doi:10.3102/0002831210371497
- Gray, J. J. (2010). *Are principals good at identifying effective teachers? A comparison of teachers' principal ratings and residual gain on standardized tests*. (Doctoral dissertation) Retrieved from Proquest Dissertations and Theses. (305206320)
- Guernsey, L., & Ochsorn, S. (2011). *Watching teachers work. Using observation tools to promote effective teaching in the early years and early grades*. Washington, DC: New America Foundation
- Hamre, B.K., Hatfield, B., Pianta, R.C., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*, 85, 1257-1274. doi: 10.1111/cdev.12184
- Hamre, B. K., Pianta, R.C., Downer, J. T., DeCoster, J., Mashburn, A.J., Jones, S., ... Hamagami, A. (2013). Teaching through interactions: testing a developmental framework of effective teaching in over 4,000 classrooms. *Elementary School Journal*, 13, 461-487. doi: 0013-5984/2013/11304-000
- Harris, D. N., & Sass, T. R. (2009). *What makes for a good teacher and who can tell?* CALDER working paper No. 30. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.
- Howes, C., Pianta, R., Bryant, D., Hamre, B., Downer, J., & Soliday-Hong, S. (2008). *Ensuring effective teaching in early childhood education through linked professional development systems, quality rating systems and state competencies: The role of research in an evidence-driven system* (pp. 1-16). Arlington, VA: National Center for Research on Early Childhood Education
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64-36. doi: IO.I037//I082-9S9X.5 1,64
- Hoyt, W. T. & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4, 403-434. doi: 1082-989X/99/S3.00
- Hsieh, W., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009) Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly*, 24, 229-247

- Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L., & Burkhauser, M. (2011). *Coaching in early care and education programs and Quality Rating and Improvement Systems (QRIS): Identifying promising features*. Report produced for the Children's Services Council of Palm Beach County. Washington, DC: Child Trends.
- Jacob, B. A. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26, 101-136. doi: 10.1086/522974
- Judge, T. A., & Ferris, G. A. (1993). Social context of performance evaluation decisions. *Academy of Management Journal*, 36, 80-105. doi:10.2307/256513
- Kline, T. J. B., & Sulsky, L. M. (2009). Measurement and assessment issues in performance appraisal. *Canadian Psychology*, 50, 168-171. doi: 10.1037/a0015668
- Kyriakides, L. (2005). Drawing from teacher effectiveness research and research into teacher interpersonal behaviour to establish a teacher evaluation system: A study on the use of student ratings to evaluate teacher behaviour. *Journal of Classroom Interaction*, 40, 2, 44-66. Retrieved from <http://www.jciuh.org/>
- La Paro, K. M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., ... Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development*. 20, 657-692. DOI:10.1080/10409280802541965
- Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, 73, 67-85. doi: 10.1348/096317900166886
- Li, Y., Hughes, J. N., Kwok, O. & Hsu, H. (2012). Evidence of convergent and discriminant validity of child, teacher, and peer reports of teacher-student support. *Psychological Assessments*, 24, 54-65. doi: 10.1037/a0024481
- LoCasale-Crouch, J., & Hamre, B. K. (2008a). *Teacher Knowledge and Skills Survey*. Unpublished manuscript, Center for the Advanced Study of Teaching and Learning, University of Virginia.
- LoCasale-Crouch, J., & Hamre, B. K. (2008b). *Teacher-Coach Relationship Scale*. Unpublished manuscript, Center for the Advanced Study of Teaching and Learning, University of Virginia.
- Mashburn, A. J., Downer, J. T., Hamre, B. K., Justice, L. M. & Pianta, R. C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science*, 14, 179-196. doi:10.1080/10888691.2010.516187
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... Early, D. M. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749. doi: 10.1111/j.1467-8624.2008.01154.x
- National Center on Quality Teaching and Learning (2012). *What do we know about coaching?* Retrieved from <http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/What-Do-We-Know-About-Coaching.pdf>
- National Research Council. (2008). *Early Childhood Assessment: Why, What, and How?* Washington, DC: The National Academies Press.
- Neuberger, J. (2012). Benefits of a teacher and coach collaboration: A case study. *The Journal of Mathematical Behavior*, 21, 290-311. doi:10.1016/j.jmathb.2011.12.004
- Noonan, L. E., & Sulsky, L. M. (2001). Impact of frame-of-reference and behavioral observation training on alternative training effectiveness criteria in a Canadian military sample. *Human Performance*, 14, 3-26. doi: 10.1207/S15327043HUP1401_02
- Office of Head Start (2013). *Use of Classroom Assessment Scoring System (CLASS™) in Head Start*. Retrieved from <http://eclkc.ohs.acf.hhs.gov/hslc/sr/quality/class>
- Ostrander, L. R. (1996). *Multiple Judges of Teacher Effectiveness: Comparing Teacher Self-Assessments with the Perception of Principals, Students and Parents*. Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12). Retrieved from <http://eric.ed.gov/?id=ED399267>
- Pianta, R.C. (2006). *Standardized Classroom Observations from Pre-K to 3rd Grade: A Mechanism for Improving Access to Consistently High Quality Classroom Experiences and Practices During the P-3 Years*. Discussion paper 103, Early Childhood Research Collaborative. Retrieved from http://humancapitalrc.org/dp_catalog/DP103.pdf
- Pianta, R.C. & Hamre, B.K. (2009). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity. *Educational Researcher*, 38, 109-119. doi: 10.3102/0013189X09332374

- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS), Manual, Pre-K*. Baltimore: Paul H. Brookes.
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K. & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 23, 431-451. doi:10.1016/j.ecresq.2008.02.001
- Potvin, G., Hazari, Z., Tai, R. H. & Sandler, P. M. (2009). Unraveling bias from student evaluations of their high school science teachers. *Science Education*, 93, 827-845. doi: 10.1002/sce.20332
- Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M.J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, 102, 299-312. doi: 10.1037/a0017763
- Raver, C.C., Jones, S. M., Li-Grining, C.P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 23, 10-26. doi:10.1016/j.ecresq.2007.09.001
- Riley, D. A., & Roach, M. A. (2006). Helping teachers grow: Toward theory and practice of an “emergent curriculum” model of staff development. *Early Childhood Education Journal*, 33, 363-370. doi: 10.1007/s10643-006-0068-0
- Roch, S. G., & O’Sullivan, B. J. (2003). Frame of reference rater training issues: recall, time and behavior observation training. *International Journal of Training and Development*, 7, 93-107. doi: 10.1111/1468-2419.00174
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-k programs predict children’s learning? *Science*, 341, 845-846. doi: 10.1126/science.1233517
- Skretta, J. (2007). Using walk-throughs to gather data for school improvement. *Principal Leadership*, 7, 16-23. Retrieved from <http://www.nassp.org/tabid/2043/default.aspx>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. New York, NY: The New Teacher Project.
- Zaslow, M., Tout, K., & Halle, T. (2011). Differing purposes for measuring quality in early childhood settings: Aligning purpose with procedures. In M. Zaslow, I. Martinez-Beck, K. Tout & T. Halle (Eds.), *Quality measurement in early childhood settings* (389-410). Falls Church, VA: Paul H. Brookes Publishing Co.
- Zaslow, M., Tout, K., Halle, T., & Forry, N. (2009). *Multiple purposes for measuring quality in early childhood settings: Implications for collecting and communicating information on quality*. OPRE Issue Brief. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, and Child Trends.
- Zaslow, M., Tout, K., & Martinez-Beck, I. (2010). *Measuring the quality of early care and education programs at the intersection of research, policy, and practice*. OPRE Research-to-Policy, Research-to-Practice Brief. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.