Assessing the Effectiveness and Interactions of Instructional Coaches: Initial Psychometric Evidence for *the Instructional Coaching Assessments–Teacher Forms*

Linda A. Reddy, PhD[1], **Todd Glover, PhD**[1] iD, **Alexander Kurz, PhD**[2], **and Stephen N. Elliott, PhD**[2]

Abstract

The conceptual foundation and initial psychometric evidence are provided for the *Instructional Coaching Rating Scales* and *Interaction Style Scales–Teacher Forms*. These forms are part of a multicomponent online assessment system designed to evaluate the effectiveness of coaching skills and interactions that support the needs of teachers and students. Specifically, the article presents the theory, evidence, and measurement framework for the system. Findings indicate that the *Rating Scales* and Interaction Style Scales–Teacher Forms have very good internal structures based on multiple fit statistics for confirmatory factor analyses, high internal consistency, good item-to-scale total correlations, and freedom from item bias. Collectively, this promising statistical evidence is supportive of valid score inferences. Study limitations and directions for research are discussed.

The implementation of Every Student Succeeds Act (ESSA, 2015) has placed greater responsibility on state education departments for effective educator evaluation and professional development (PD). As a result, states are implementing changes to enhance PD that can improve classroom instruction and result in greater student achievement and overall school improvement (Mead, 2012; Reddy, Kettler, & Kurz, 2015; Sawchuk, 2016).

Instructional coaching, a job-embedded PD approach, is a means of overcoming the limitations of workshop-based PD with respect to the transfer of knowledge and skills into classroom practices (Gulamhussein, 2013). Despite the increased popularity of instructional coaching in schools, research is needed to better operationalize, assess, and evaluate the implementation of essential coaching components (Denton & Hasbrouck, 2009; Glover & Reddy, 2017). The purpose of the research reported is to provide initial psychometric evidence for a multicomponent system for assessing research-based skills and interactions of instructional coaching.

Instructional Coaching to Enhance PD and Improve Student Achievement

Most scholars agree that coaching is a job-embedded, individualized, data-driven, and sustained practice (e.g., Denton

& Hasbrouck, 2009; Glover & Reddy, 2017). Although approaches to school-based coaching vary in scope and foci, many assert coaching models share a focus on prioritizing instructional needs, established goal-driven plans of support, modeling, facilitating teacher practice, and providing ongoing regular feedback to promote high-fidelity instructional practices (Denton & Hasbrouck, 2009; Erchul, 2015; Kurz, Reddy, & Glover, 2017; Showers, Joyce, & Bennett, 1987). Coaching is distinct from consultation in that coaching is primarily designed to support teachers' continuous PD. Many school-based coaches were former effective classroom teachers, whereas consultants tend to have specialized training and experiences in psychology, business, and/or other allied fields (e.g., occupational therapy, speech and language, and behavior interventions). Often coaches reside within the district or school with supervision from an administrative leader (i.e., curriculum

[1]Rutgers University, Piscataway, NJ, USA
[2]Arizona State University, Tempe, USA

Corresponding Author:
Linda A. Reddy, Graduate School of Applied and Professional Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA.
Email: LReddy@gsapp.rutgers.edu

director and assistant superintendent). Given the distinct role of coaches in schools, ongoing evaluation of key coaching competencies via technically sound assessments is essential to enhancing effective practices.

Although, as outlined in the ESSA (2015), instructional coaching is important for maximizing effective instruction, there are few research-based coaching models and even fewer coaching measures and resources available to assess and develop coaching skills and interactions. Most coaching approaches and assessments are also narrow in scope, focusing on content-specific practices such as literacy/reading, mathematics, and science (e.g., American Institutes for Research [AIR], 2005; APQC Education Advanced Working Group, 2011). The effectiveness of coaching also varies widely, often due to a lack of adherence to a defined coaching process and a clear focus on research-based coaching actions and outcomes that guide PD (e.g., Denton & Hasbrouck, 2009; Glover & Reddy, 2017). For example, without explicit coaching guidelines and assessment-driven performance feedback, coaches in the Reading First Program from the No Child Left Behind Act of 2001 implemented a range of loosely defined activities to support teachers across the nation, resulting in suboptimal teacher and student outcomes (AIR, 2005; Deussen, Coskie, Robinson, & Autio, 2007). Thus, when the process of coaching is poorly defined, assessed, and monitored over time, the consequences often are not positive, resulting in missed opportunities to improve instruction that enhances student achievement and school effectiveness.

Current evaluation practices of instructional coaches rely on tools and standards meant for classroom teachers. This is a significant problem because the process, actions, and intended outcomes for classroom teaching versus instructional coaching are different. School leaders often charged with evaluating coaching staff must decide what teacher standards to use when making effectiveness decisions and PD recommendations. This can lead to inconsistent coaching evaluation practices and poorly defined feedback for PD (Killion, Harrison, Bryan, & Clifton, 2012). Thus, providing instructional coaches skill-focused and valid assessment-based performance feedback is critical to their development and effectiveness in fostering teacher PD (Reddy, Dudek, & Lekwa 2017).

Scholars have reported that the provision of formative feedback can effectively modify learners' knowledge, thinking, or behavior that lead to improved learning (e.g., Brophy, 1981; Hagermoser Sanetti, Luiselli, & Handler, 2007; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Schwartz & White, 2000; Shute, 2007). Unfortunately, the availability of coaching assessments is sparse. A comprehensive literature review of published or publicly available coaching assessment tools worldwide between 1980 and 2016 (Reddy, 2016) resulted in a total of 26 assessments, with 25 assessments publicly posted with no validity evidence and one

published article that described the content validity of two Literacy Coach Appraisal Instruments (Lane, Robbins, & Price, 2013). Specifically, these assessments utilized a 4-point Likert-type scale measuring eight broad domains of coaching (e.g., Curriculum, Teachers; Staff Development; Technology; Liaison; Assessment; Home and Family Education; Resource Management; Professional Disposition). The Literacy Coach Appraisal Instruments, however, do not have evidence of reliability and construct validity and do not offer multirater assessment. In summary, none of the 26 assessments measure content-neutral, core process-focused coaching skills and interactions, nor do they offer multirater, online applications. As with teachers, there is an ongoing need to evaluate coaches in a variety of contexts supporting teachers in various instructional areas (Denton & Hasbrouck, 2009; Glover & Reddy, 2017; Kurz et al., 2017; Showers et al., 1987). Despite the range of coaching practices, the literature in education, sports, and executive coaching business points to core process-focused skills and interactions that drive coaching effectiveness (e.g., Denton & Hasbrouck, 2009; Kurz et al., 2017; Showers et al., 1987). These core process-focused skills and interactions are important for advancing teacher practice and student outcomes and can be evaluated via a content-neutral assessment approach. Teacher assessments of coaches' performance are essential for improving the key skills and interactions of coaches and the overall process of coaching for the PD of teachers. This study was the initial investigation of the utility and validity of the *Instructional Coaching Assessments–Teacher Forms*. Subsequent research is underway with the two remaining Assessment forms—supervisors and coaches themselves—in the 360 or comprehensive multirater instructional coaching system.

The Development of an Assessment Designed to Improve Instructional Coaching

The Instructional Coaching Assessment (Reddy, Glover, Kurz, & Elliott, 2017) is an online, multirater assessment system that provides feedback reports to support the evaluation and development of instructional coaching talent. The assessment approach involves conducting a 360° assessment, completed by the coach, teachers served by the coach, and/or the coach's supervisor. A 360° assessment offers a comprehensive assessment of coaching effectiveness and interactions by capturing feedback from key stakeholders involved in the coaching process. Each stakeholder provides unique and complimentary perspectives on the coaching process that is valuable for informing PD for coaches. This study examined the initial psychometric properties of the *Rating Scales* (RS) *and Interaction Style Scales* (ISS)—*Teacher Forms*, the cornerstones for responsive instructional coaching.

Theoretical and Empirical Foundations of the Instructional Coaching Assessments

Grounded in the behavior consultation and coaching literature (Bergan & Kratochwill, 1990; Bergan, 1977; Kurz et al., 2017; Sheridan & Kratochwill, 2008), the instructional coaching assessments were developed to evaluate the effective implementation of core coaching skills or actions and coaching interactions with teachers via a content-neutral, process-focused coaching framework. This literature highlights key consultant/coaching actions that yield moderate to large effects on multiple school stakeholders (i.e., classroom teachers, special educators, school psychologists, content area specialists, and parents) and student academic and behavior outcomes (e.g., Erchul & Sheridan, 2014; Reddy, Barboza-Whitehead, Files, & Rubel, 2000; Sheridan et al., 2012). In addition, rigorous evaluations of coaching efficacy on early reading (Glover, 2017) and teacher universal practices and student academic behavior further support specific core coaching skills (Fabiano, Reddy, & Dudek, 2018; Reddy, Dudek, & Lekwa, 2017) and offer further insights into universal coaching skills. These coaching skills/actions include identifying teacher and student needs with data, setting specific measureable goals, designing interventions and supports matched to needs/goals, teaching and modeling of intervention steps, providing practice opportunities, delivering ongoing performance feedback, and evaluating intervention fidelity and progress toward goals. These skills/actions have been found to facilitate consultation and coaching outcomes (Kurz et al., 2017; Reddy et al., 2000) and were used to develop the scales and items of the assessments (described next).

In a review of the education, sports, and executive coaching literature, Kurz et al. (2017) proposed a multidisciplinary framework that highlights the scope of coaching. Specifically, they suggested coaching should have one of the three foci: (a) skills: coaching targets discrete skills (e.g., support for a teacher's use of positive behavioral praise with her students); (b) process: coaching targets a process or progression of activities (e.g., support for the use of a multiphase data-based decision-making model); or (c) development: coaching targets the application of skills and processes to achieve growth toward professional goals. Based on this framework, the assessments are designed to address the skills, process, and development of coaching.

Collectively, the theoretical and empirical literature has informed the development of four interrelated assessments—RS, Performance Monitoring Rubrics (PMRs), ISS, and Observation Record (OR). This complementary set of assessments generates results for guiding improvements in the instructional coaching process based on three interdependent phases (Goal Formulation > Implementation Support > Intervention Evaluation) that feature six coaching actions (Identifying Needs and Resources, Setting Goals, Designing Implementation Plans, Modeling Implementation Steps, Providing Performance Feedback, and Evaluating Implementation and Goal Attainment). As shown in Figures 1 and 2, the six coaching actions in the three phases intersect with three outcomes (i.e., scales: Quality Instruction, Positive Behavior Management, and Responsive Learning Communities).

The construct and item development of the instructional coaching assessments were guided by classical test theory, established psychometric standards (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 2014), and recent evidence-centered design principles (Mislevy, Almond, & Lukas, 2003). Evidence-centered assessment design is an approach to developing educational assessments based on evidentiary arguments (Mislevy et al., 2003). Evidence-centered assessments generate aggregate information from various sources and provide evidence to support score inferences that individuals possess specific skills and competencies. The development of items and scales was informed by several methods: (a) comprehensive review of scholarly, peer-reviewed publications in the fields of education, sports, and executive coaching; (b) critique of other related scales; (c) educational assessment and instructional scholars' feedback; (d) consumer feedback; (e) school-based field testing; and (f) a series of construct and item analyses.

The central assessment component of the system is a skill-focused behavior rating scale (see Figures 1 and 2). The RS are centered on three coaching outcomes (i.e., scales: Quality Instruction, Positive Behavior Management, and Responsive Learning Communities). Each scale includes three phases (Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills), with two coaching actions nested within each process phase. The assessment system is designed to measure specific coaching skills and actions that lead to improved teacher outcomes (i.e., enhanced Quality Instruction, Behavior Management, and Responsive Learning Communities) and are related to positive student learning and behavior (Erchul & Sheridan, 2014; Reddy et al., 2000; Sheridan et al., 2012).

As an evidence-centered assessment, the measures are designed to generate data-specific performance feedback (scores) that assesses coaching skill-focused actions to promote positive coaching outcomes/competencies. The Instructional Coaching assessments provide a comprehensive evaluation of a coach's effectiveness at implementing problem-solving actions with teachers. This involves the collection of sources of evidence that support performance feedback (scores) from multiple informants (Shute, Kim, & Razzouk, 2013). The evidence-centered, action–outcome framework provides a systematic approach to measure and drive continuous improvement for coaching talent and schools.

Scales	The Instructional Coaching Scales[a]
Coaching Outcomes	**Quality Instruction:** Data-based instructional decisions and practices that maximize students' academic performance. Instructional and assessment practices are systematically aligned with expectations and benchmarks for promoting student academic performance.
	Positive Behavior Management: Data-based behavior management decisions and practices that improve social interactions and healthy participation in the classroom and school. Behavior management and assessment practices are systematically aligned with expectations and rules for promoting positive behavior.
	Responsive Learning Communities: The identification and systematic sharing of information and resources address needs and goals of communities of teachers and students. Communities are present that promote collaboration, professional development, and a data-based culture that maximize student learning.
Phases	**The Instructional Coaching Actions**
Goal Formulation Skills	**Identifying Needs and Resources:** Coaches communicate and work collaboratively with teachers to gather information to identify needs and resources to improve student performance and teacher practices. This information may guide the formulation of goals for quality instruction, behavior management and school-wide learning and professional development.
	Setting Goals: Coaches collaborate with teachers to identify and write measureable goals based on student functioning and teacher practices relative to performance expectations. Measureable goals are specific, reflect high expectations and are essential for designing effective implementation plans.
Implementation Support Skills	**Designing Implementation Plans:** Coaches collaborate with teachers to create specific steps needed to achieve goals. Coaches consider teachers' skills, possible resources and barriers to ensure plans can be modeled and delivered with integrity.
	Modeling Implementation Steps: Coaches demonstrate steps specified in implementation plans and support teacher practice of steps. Modeling and practice help teachers to independently deliver a plan with integrity and create opportunities for feedback.
Evaluation Skills	**Providing Performance Feedback:** Coaches use data to provide specific, positive, and timely feedback about plan implementation and goal attainment. Feedback helps refine implementation of evidence-based practices that improve student and teacher performance.
	Evaluating Plan Implementation and Goal Attainment: Coaches use a comprehensive set of data to make summative judgments about the integrity of plan implementation and progress toward goal attainment. This information is used to continuously improve plans, goal attainment, and coaching effectiveness.

Figure 1. The assessment action—outcome framework and definitions.
[a]Each rating scale includes a total of 30 items, each phase includes 10 items, and each action includes five items.

The assessments measure instructional coaching as a problem-solving, interpersonal process aimed at enhancing educators' instructional and classroom management practices to maximize achievement for all students. Within this framework, instructional coaching theoretically improves school-wide effectiveness through support for both individual teachers and data-based professional learning communities. In turn, student outcomes are improved via enhanced classroom and school-wide practices. The assessments were designed specifically to collect and integrate information as feedback about (a) a coach's relative strengths and areas for development, (b) a coach's development of competencies, (c) the overall effectiveness of a coach at implementing actions known to influence problem solving, and (d) PD of coaches that likely will lead to school improvement.

Assessment components. Assessments include four online assessments and feedback reports to inform continuous improvement for coaching and continuing PD for teachers.

All ratings are based on documented sources of evidence from recent coaching interactions to ensure ratings are meaningful and based on evidence. Feedback reports provide a summary of the assessment results organized around six guiding questions: *Who evaluated me and what evidence did they use to determine my effectiveness? What were my overall coaching effectiveness ratings? Which specific coaching skills were identified as strengths and which skills could be improved? How were my coaching interaction characteristics perceived? Which characteristics were considered in need of improvement? How was my observed coaching performance evaluated?*

The RS are completed by the coach, teachers served by the coach, and the coach's supervisor to evaluate 30 items pertaining to six actions within each of the three (separate) outcome areas/scales: Quality Instruction, Positive Behavior Management, and Responsive Learning Communities (see Figure 1). The RS can be used for formative or summative assessment. Scores are based on a 5-point Effectiveness

	Coaching Subscales and Actions					
	Goal Formulation Skills		Implementation Support Skills		Evaluation Skills	
Coaching Outcomes (Scales)	Identifying Needs and Resources	Setting Goals	Designing Implementation Plan	Modeling Implementation Steps	Providing Performance Feedback	Evaluating Implementation and Goal Attainment
Quality Instruction	Action Score	Action Score	Action Score	Action Score	Action Score	Action Score
Positive Behavior Management	Action Score	Action Score	Action Score	Action Score	Action Score	Action Score
Responsive Learning Communities	Action Score	Action Score	Action Score	Action Score	Action Score	Action Score
Subscale Scores	Subscale 1 Score		Subscale 2 Score		Subscale 3 Score	
Total Score	Coach Total Score {Subscales 1 + 2 + 3 Summed}					

Figure 2. Instructional Coaching measurement framework—Rating Scales.

rating scale (Ineffective to Exceptionally Effective) and are transformed to Total Scale scores and subscale scores of Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills (10 items each) and six skill-focused action scores (five items each; see Figure 2). Scores are used to guide coach improvement. For the Quality Instruction Scale, the coaching action of Identifying Needs and Resources includes items such as *Listening to teachers describes instructional approaches and resources* and *Using data (e.g., observations and lesson plans) to inform needs for instructional improvement*. For the Positive Behavior Management Scale, the coaching action of Identifying Needs and Resources includes items such as *Assessing the match between behavior supports and student behavior* and *Observing behavior management practices and use of resources needs*. For the Responsive Learning Communities Scale, the coaching action of Identifying Needs and Resources includes items such *Synthesizing data across groups to identify needs for improving practices* and *Communicating information and resources for improving practices and student learning*. This study presents psychometric findings on the RS–Teacher Form.

PMRs, a derivative of the RS, are completed by the coach and teachers served by the coach to briefly assess coaching effectiveness for various coaching actions within each of the three phases of coaching (i.e., formulating goals, designing and implementing plans, and evaluating the delivery of plans and progress toward goals) throughout a given case. That is, both a coach and the teacher involved in a case independently complete an action-specific PMR (e.g., Setting Goals, Designing Implementation Plan, and Providing Performance Feedback) immediately after the completion of that action for the purpose of enhancing communication and refining the coaching process so that it is likely to be more effective for the teacher. These rubrics are based on high-loading items from the RS subscales. The PMR utilizes a five-level rubric of effectiveness (Ineffective to Exceptionally Effective) to assess each of the six coaching actions as they relate to each of the coaching actions (Quality Instruction, Behavior Management, and Responsive Learning Communities). Note that research to establish initial psychometrics of these PMRs is underway; only findings on the perceived usability of the PMR by coaches and teachers is presented in this report.

The ISS are completed by the coach, teachers served by the coach, and the coach's supervisor to evaluate key interpersonal coaching characteristics that can support or hinder collaboration. The ISS can be used for formative or summative assessment. The scale contains 28 items that factor into two scales: Positive Coaching Characteristics (PCC; e.g., *a good listener, skills*, and *reliable*) and Negative Coaching Characteristics (NCC; e.g., *disorganized, unpredictable*, and *abrupt*). Each item on these scales is rated on two dimensions: Prevalence (rated on a 5-point scale from *Not Prevalent* to *Highly Prevalent*) and Need for Improvement (yes or no). The ISS yields scores for both Positive Coaching Characteristics and Negative Coaching Characteristics scales and endorsed items Needing Improvement. This study presents psychometric findings on the ISS–Teacher Form.

The OR is a summative multicomponent rubric for assessing coaching behaviors based on key sources of evidence pertaining to the six skill-focused coaching actions. It is used by supervisors to evaluate coaches and make recommendations for coaching improvements. Supervisors read quality
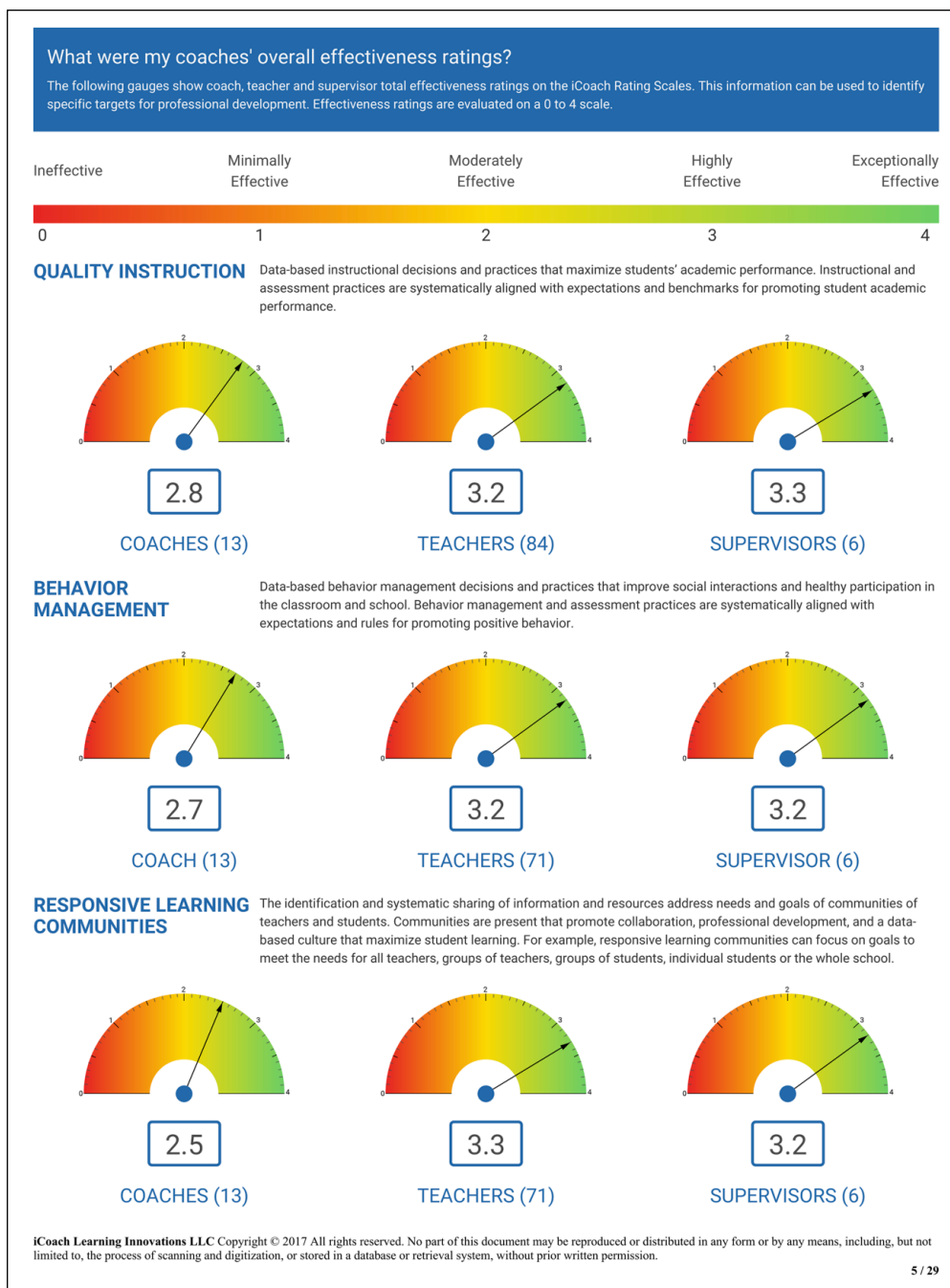
Figure 3. Instructional Coaching rating scale district feedback report example—overall effectiveness.

indicator statements provided within a five-level rubric to determine the level that best describes current effectiveness (Ineffective to Exceptionally Effective) of each of the six skill-focused coaching actions. Note that research to establish initial psychometrics of the OR is underway.

Feedback reports. The assessments generate on-demand, integrated data-based feedback that facilitates a process of continuous improvement whereby coaches guide teachers and school teams through the six coaching actions for improving outcomes of instruction, behavior management, or responsive learning communities. For example, coaches are provided comparative feedback from themselves, their teachers, and supervisors on their overall effectiveness in promoting instruction, behavior management, and responsive learning communities (see Figure 3). Coaches are also provided feedback on actions that are rated as strengths and those actions rated as needing improvement (see Figure 4).
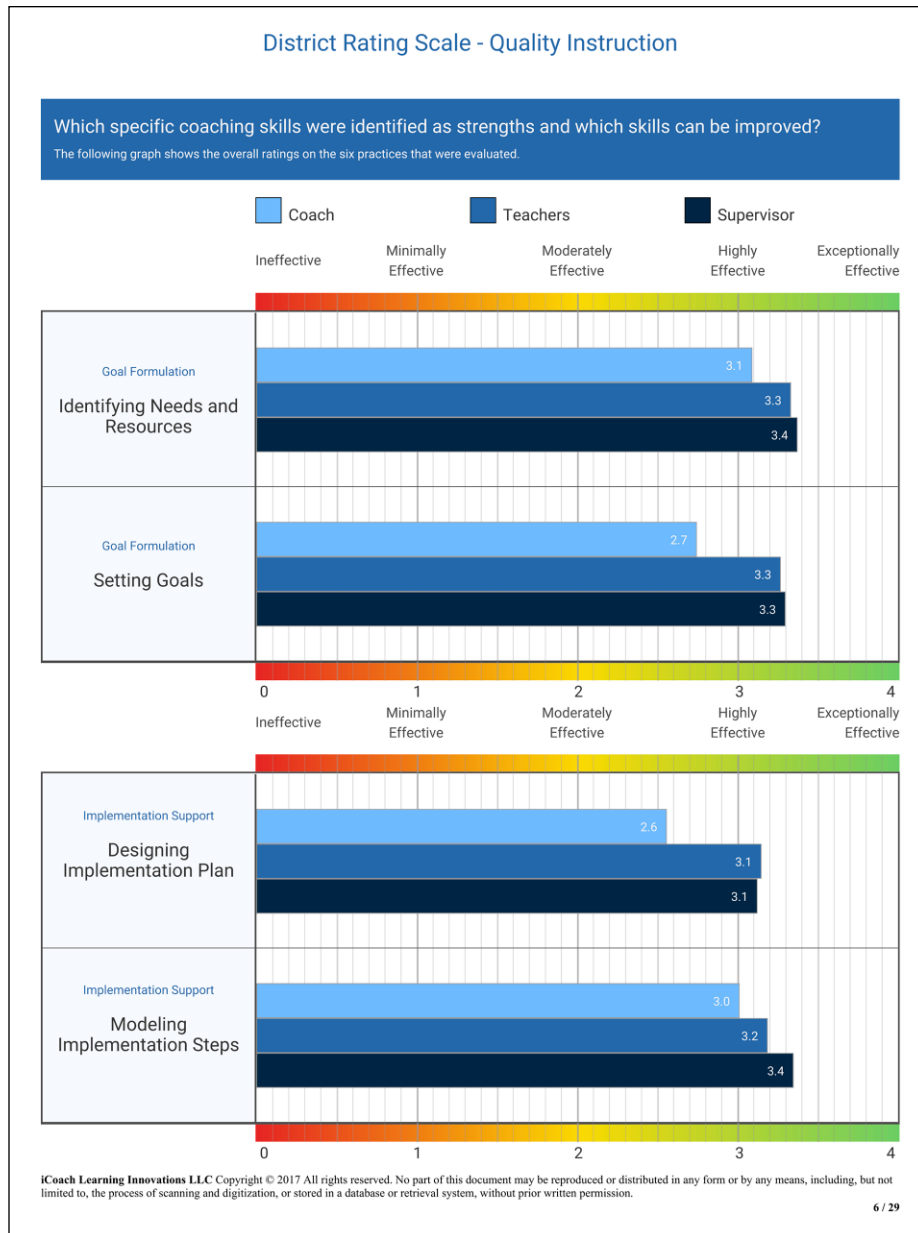
Figure 4. Instructional Coaching rating scale district feedback report example—coaching actions.

The feedback reports provide valuable information for creating targeted PD for coaches and monitoring coaching improvements over time. In addition, aggregate reporting and graphic performance feedback are available for three descriptive groups/levels: school district, school, and individual coaches. A number of utility analyses have been conducted with the Feedback Reports to ensure they communicate efficiently and effectively with users. An expert group of four technology developers reviewed and provided feedback on the report content, layout, clarity, and use. In addition, two consumer groups consisting of 46 school-based instructional coaches and seven supervisors of

coaches from two states reviewed and provided input on the usability of the feedback reports.

Research Questions and Predictions

The primary purpose of this study was to establish initial evidence for reliability and validity of the score inferences from the RS and ISS–Teacher Forms and the preceived usability of the PMR by coaches and teachers. These assessments have the potential to fill a void by offering a technology-supported action–outcome framework that enables school personnel to systematically evaluate coaches and

advance coaching interactions that support teachers' continuing PD. In this study, we addressed four questions:

Research Question 1: What is the internal structure of the RS–Teacher Form, as measured by internal consistency, item-to-total correlations, and confirmatory factor analysis (CFA)?

Research Question 2: What is the internal structure of the ISS–Teacher Form, as measured by internal consistency, item-to-total correlations, and exploratory factor analysis?

Research Question 3: Are the RS and ISS–Teacher Forms free from item bias with respect to common teacher demographics?

Research Question 4: What is the perceived usability of the PMR by coaches and teachers who have been involved in instructional coaching?

We expected that the theorized internal structure of the RS and ISS for teachers would be confirmed empirically and meet acceptable technical benchmarks (Cicchetti, 1994; Jackson, Gillaspy & Purc-Stephenson, 2009). We also expected the items would function similarly across teacher subgroups (freedom from item bias). Finally, we expected users of the PMR to rate the PMR scales favorable on a common usability scale (Bangor, Kortum, & Miller, 2008).

Method

Sample

The sample included 25 full-time instructional coaches who coached approximately 225 general and special education teachers. Kindergarten through 12th grade teachers from 11 high poverty schools participated in a statewide school reform grant funded by the U.S. DOE Teacher Incentive Fund (Reddy, Kettler & Kurz; 2012–2018). Nearly 75% of students enrolled in these schools received free or reduced lunch.

Coaches were assigned to schools by their school administrators. Approximately 12 teachers were served by each coach. Teachers were coached by a single coach during the school year. The coaches were predominantly Caucasian (65%) females (100%) with an average age of 46 years (*SD* = 10.20); 70% had Masters degree and on average had 15 years (*SD* = 7.2) of teaching experience. Coaches were former classroom teachers who were identified as effective teachers based on a multimethod educator evaluation system aligned with state policy.

The teacher sample (*n* = 225) included predominantly Caucasian (66%) females (78%) with an average age of 32 years (*SD* = 9.27). The average number of students

per classroom was 21 (*SD* = 3.94). Teacher educational backgrounds included 65% with a bachelor's degree. The average number of years of teaching experience was 5.44 (*SD* = 5.49). In the participating schools, some teachers taught multiple grades. Thus, teacher grade bands included approximately 68 teachers assigned to kindergarten to Grade 2, 48 assigned to Grades 3 to 5, 20 assigned to K-5 grades, 32 assigned to Grades 6 to 8, and 57 assigned to Grades 9 to 12.

Instruments

RS. The RS–Teacher Form includes three scales (Quality Instruction, Positive Behavior Management, and Responsive Learning Communities; see Figure 1). Each RS includes a Total scale of 30 items, three subscales (i.e., Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills; each includes 10 items), and six action scores (five items each; see Figures 1 and 2). Teachers indicate the sources of evidence (i.e., personal interactions, written records, presentations and materials, and observations) used to evaluate his or her coach for each RS subscale. Teachers are then asked to rate the effectiveness of their coach's actions on a 5-point Likert-type item scale, with 0 = *ineffective*, 1 = *minimally effective*, 2 = *moderately effective*, 3 = *highly effective*, and 4 = *exceptionally Effective* (see Figures 1 and 2). Item ratings for each of the six coaching actions are totaled and divided by 5 to yield a mean item raw score that is on the 5-point effectiveness continuum.

ISS. Teacher Form includes 28 items and two scales— PCC and NCC—rated on a 5-point Likert-type scale of Prevalence (i.e., 0 = *not prevalent*, 1 = *minimally prevalent*, 2 = *somewhat prevalent*, 3= *prevalent*, and 4 = *highly prevalent*). Items are also rated on Need for Improvement (i.e., yes or no). The PCC scale includes 20 items and the NCC scale includes eight items. Prevalence ratings for the PCC are totaled and divided by 20 to yield a mean item raw PCC score, whereas the prevalence ratings on the NCC are totaled and divided by 8 to yield a mean item raw NCC score; both prevalence mean item scores are reported on the 5-point prevalence scale. The Need for Improvement items are used to signal characteristics that raters believe deserve improvement and do not yield a total score. For purposes of this study, we have not reported on this score.

PMR. The PMR includes six performance coaching action areas that follow three phases: Goal Formulation, Instructional Support, and Evaluation. A definition and effectiveness level descriptions are provided for each skill performance area. Each performance action area is characterized by five

levels of effectiveness (i.e., *Exceptionally Effective, Highly Effective, Moderately Effective, Minimally Effective*, or *Ineffective*; see Table 2 example). The raters—the coach and the teacher involved in problem solving—independently select an effectiveness level that best describes the current level of coaching performance for work on your current case.

System Usability Scale *(SUS).* To determine the usability of the PMR, we adapted the SUS by Brooke (1996) and its refined version by Bangor, Kortum, and Miller (2008). The 5-point Likert-type scale features 10 items (0 = *strongly disagree* to 4 = *strongly agree*), which are multiplied by 2.5 to arrive at a total score range of 0 to 100. The authors report that SUS Cronbach's alphas were consistently reported above .90 and correlated relatively high with other usability scales (e.g., .84 with the Usability Metric for User Experience). Moreover, normative and criterion-referenced interpretations of the total score are also available (Bangor et al., 2008; Finstad, 2010). Based on more than 200 surveys, the mean total score for the surveys was 70.14, with an *SD* of 21.71. Our scale included 10 questions with a 5-point Likert-type scale scored in the same way as the SUS (see items in Table 3).

Procedures

School-based instructional coaches were trained on the instructional coaching multiphase, data-driven coaching model. Training consisted of three full days prior to the start of school year (2015–2016) and eight, monthly full-day implementation support trainings during the school year (approximately 55 hr of coaching training/support). Teachers were provided brief in-person training and written information on the coaching process prior to enrollment. Data collection was conducted in the spring of 2016 via an online application. Participating teachers were provided one-on-one school-based instructional coaching that focused on improving teacher practices and student learning or behavior needs. Coaching was job-embedded, collaborative, data-driven, goal-focused, and sustained over time. Coaches provided modeling, practice, and performance feedback. Coaches used data to inform decisions on identifying needs, set goals, designed intervention/support plans, provided ongoing feedback based on classroom observation and other data sources, and evaluated implementation fidelity and progress toward goals. On average, teachers set two goals/targets that generally fell under teacher practice (e.g., increase the frequency and quality of teachers' academic performance feedback, praise for appropriate behavior, and direct instructional methods) or student outcomes (e.g., increase oral reading fluency and accuracy, and student on-task behavior or reduce disruptive/

aggressive behaviors). On average, teachers received six coaching sessions and seven classroom observations during the school year. Overall, teacher reported high satisfaction with coaching supports.

Data Analysis

To address the four research questions of interest, a set of descriptive and psychometric analyses were conducted. First, descriptive statistics with the RS and ISS–Teacher Forms for the entire sample were completed. Second, CFA was used to test the three RS model fit to the data and exploratory factor analysis was used to examine the internal structure of the ISS using technical benchmarks. Third, we computed coefficient alphas, standard error of measurement (SEM), and item-to-scale correlations for the RS Total and subscales, as well as ISS scales. Fourth, Reynolds and Carson's (2005) partial correlation model for detection of differential item functioning (DIF) was used to identify potential bias in the RS and ISS–Teacher Forms items for teacher age, educational degree, and years of teaching experience. The partial correlational method has been used in numerous publications since the 1984 original publication (e.g., Reddy et al., 2013a; Reddy et al., 2015; Reddy, Dudek, & Shernoff, 2016; Reynolds & Carson, 2005). In this investigation, freedom of item bias was examined using several methods, including partial correlations, as well as expert and consumer input. Finally, descriptive statistics were computed to assess coaches' and teachers' perceived utility ratings of the PMRs.

Results

Descriptive Statistics

Descriptive statistics were computed for the RS and ISS. RS mean item raw scale scores and *SD*s were as follows: Quality Instruction $M = 3.23$ ($SD = 0.83$); Positive Behavior Management $M = 3.22$ ($SD = 0.89$); and Responsive Learning Community $M = 3.14$ ($SD = 0.95$). ISS mean item raw scale scores and SDs were as follows: PCC $M = 3.56$ ($SD = 0.31$) and NCC $M = 0.45$ ($SD = 0.33$). Raters completed the various online assessments as designed, so missing data was not a problem.

Evidence to Support the Internal Structure of the Instructional Coaching Assessments

RS. For the RS–Teacher Form, CFAs and maximum likelihood estimation were computed to test the hypothesized three-factor model (Goal Formulation > Implementation Support > Intervention Evaluation) fit to the six mean action scores for each RS separately (see Tables 1 and 2) using AMOS version 23. Using the mean action scores,

Table 1. Goodness of Fit Indices for the Rating Scales: Quality Instruction, Behavior Management, and Responsive Learning Communities (Teacher Form).

Indices	Quality Instruction	Positive Behavior Management	Responsive Learning Communities
CMIN/DF	0.742	1.385	0.471
RMR	0.003	0.002	0.001
GFI	0.943	0.914	0.961
AGFI	0.800	0.698[a]	0.863
NFI	0.982	0.975	0.877[a]
TLI	1.017	0.982	2.004
CFI	1.000	0.993	1.000
RMSEA	0.000	0.127[a]	0.000
PCLOSE	0.650[a]	0.250[a]	0.850[a]

Note. Model fit was based on the following: CMIN/DF = chi-square/*df* (<2.0; Arbuckle, 2015; AMOS, p. 587); RMR = root-mean-square residual (<0.08; Hu & Bentler, 1999); GFI = goodness-of-fit index (≥0.90; Jöreskog & Sörbom, 1986); AGFI = adjusted goodness-of-fit index (≥0.80; Jöreskog & Sörbom, 1986); NFI = normed fit index (>0.90; Jöreskog & Sörbom, 1986); TLI = Tucker–Lewis index (>0.95; Jackson et al., 2009); CFI = comparative fit index (>0.95; Jackson et al., 2009); RMSEA = root mean square error of approximation (<0.06; AMOS, p. 590; Arbuckle, 2015; Hu & Bentler, 1999); and PCLOSE = *p* value for test of close fit (<0.05; AMOS, p. 590; Arbuckle, 2015).
[a]Index *below* acceptable benchmark.

the ratio of sample size to number of estimated parameters was approximately 1:5 (Bentler, 1985) providing fair power. As shown in Table 1, model fit was tested using several recommended fit indices by Jackson, Gillaspy, and Purc-Stephenson (2009) and others. Model fit was based on the following: CMIN/DF = chi-square/*df* (<2.0; AMOS, p. 587; Arbuckle, 2015); RMR = root-mean-square residual (<0.08; Hu & Bentler, 1999); GFI = goodness-of-fit index (≥0.90; Jöreskog & Sörbom, 1986); AGFI = adjusted goodness-of-fit index (≥0.80; Jöreskog & Sörbom, 1986); NFI = normed fit index (>0.90; Jöreskog & Sörbom, 1986); TLI = Tucker–Lewis index (>0.95; Jackson et al., 2009); CFI = comparative fit index (>0.95; Jackson et al., 2009); RMSEA = root mean square error of approximation (<0.06; AMOS, p. 590; Arbuckle, 2015; Hu & Bentler, 1999); and PCLOSE = *p* value for test of close fit (<0.05; AMOS, p. 590; Arbuckle, 2015). As shown in Table 1, results revealed overall good model fit to the data for each RS. Specifically, eight out of nine fit indices met acceptable benchmarks for Quality Instruction, six out of nine fit indices met acceptable benchmarks for Positive Behavior Management, and seven out of nine fit indices met acceptable benchmarks for Responsive Learning Communities.

ISS. For the ISS–Teacher Form, principal components analyses, using varimax rotation with Kaiser normalization, were carried out to test the hypothesized two-factor model of PCC and NCC scales (Meyers, Gamst, & Guarino, 2013) using item-level prevalence scores due to sample size (see Table 2). SPSS version 23.0 was used. For Factor 1, 20 Positive Characteristic items yielded item loadings ranging from .675 to .813 (representing 49.66% of the variance). For Factor 2, eight Negative Characteristic items yielded item loadings ranging from .693 to .859 (representing

Table 2. Factor Loadings for the Interaction Style Scales–Positive and Negative Characteristics Scales (Teacher Form).

Scale	Factor 1	Factor 2
Positive Coaching Characteristics Scale[a]		
Likeable	0.739[b]	−0.165
A good listener	0.777	−0.249
Positive	0.720	−0.279
Sensitive to my needs	0.813	−0.173
Direct	0.675	−0.039
Respectful	0.781	−0.297
Supportive	0.794	−0.312
Skilled	0.853	−0.095
Reliable	0.857	−0.196
Knowledgeable	0.781	−0.132
Practical	0.789	−0.092
Creative	0.743	−0.024
Confident	0.826	−0.083
Accessible	0.743	−0.098
Helpful	0.792	−0.134
Efficient	0.800	−0.056
Flexible	0.792	−0.216
Goal-oriented	0.821	−0.156
Energetic	0.797	−0.075
Responsible	0.817	−0.254
Negative Coaching Characteristics Scale[c]		
Anxious	−0.050	0.693
Negative	−0.131	0.778
Forgetful	−0.081	0.813
Disorganized	−0.075	0.742
Uncertain	−0.157	0.859
Unpredictable	−0.205	0.792
Disinterested	−0.294	0.786
Abrupt	−0.149	0.694

[a]α = 970. [b]Factor loadings of principal components analyses (varimax rotation with Kaiser normalization); Comrey and Lee's (1992) guidelines for loadings in a rotated component matrix. 1.0 to .710, *excellent*; .709 to .630, *very good*; .629 to .550, *good*; .549 to .450, *fair*; and .320 or lower, *poor*. [c]α = .955.

Table 3. Usability Ratings for PMR by Coaches and Their Teachers.

Question	Coaches[a] M (SD)	Teachers M (SD)
1. I think that I could use this product to communicate progress in the coaching process.	6.20 (2.54)	7.23 (2.32)
2. I found the product unnecessary.	5.46 (2.95)	4.27 (2.86)
3. I thought the product easy to use.	7.41 (2.64)	7.53 (2.20)
4. I would prefer my responses to the assessment be anonymous.	8.43 (2.51)	7.33 (2.88)
5. I liked the five-level rating system.	6.76 (2.28)	7.67 (2.11)
6. I think the use of the assessment will have positive consequences on the outcomes of the coaching experience.	5.56 (2.89)	7.39 (2.13)
7. I could complete the rubric associated with each coaching action right after the appropriate session to provide quick and accurate information.	5.09 (2.45)	6.91 (2.50)
8. I think this product could be intrusive into the coaching process.	5.46 (2.30)	5.27 (2.80)
9. The instructions for using the performance rubric were clear.	7.04 (3.18)	7.70 (2.07)
10. Overall, I really liked using this performance rubric to summarize my perspectives on the quality of my coaching interaction.	5.19 (2.39)	6.74 (2.51)
Total	62.59 (11.51)	68.02 (11.47)

Note. PMR = Performance Monitoring Rubrics.
[a]n =27. [b]n = 177.

19.28% of the variance), respectively. Items loaded on one factor each. Based on Comrey and Lee's (1992) interpretative guidelines for rotated component matrices, the Positive and Negative Characteristic factor item loadings were found to be *very good* (.630–.709) to *excellent* (.710–.999), providing evidence for construct validity.

Evidence for the Internal Consistency of the Instructional Coaching Assessments

The RS and ISS–Teacher Forms yielded good internal consistency (scale: Cronbach's alphas ranging from .97 to .99). Specifically, internal consistency estimates were .99 for each of the scales of RS Quality Instruction, Positive Behavior Management, and Responsive Learning Communities scales; .96 to .98 for the subscales of Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills; and .97 and .95 for the six coaching actions. Similarly, internal consistency estimates were .97 and .95 for the ISS PCC and NCC scales (see Table 2). Acceptable item-to-scale total correlations that exceeded the criterion of .70 were observed for both the RS and ISS scales (Crocker & Algina, 1986; Obermiller & Spangenberg, 1998).

Based on these internal consistency indices, we calculated the SEM for each of the scales and subscales. For the RS, the SEMs for the Quality Instruction, Positive Behavior Management, and Responsive Learning Communities total scales were similar and low, ranging from .08 to .09. Similarly, the RS subscales of Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills were similar and low, ranging from .12 to .15. For the ISS, PCC and NCC scales, SEMs also were very low (i.e., .06). These relative small SEMs suggest good overall precision among the RS and ISS scales and subscales.

Evidence for the Freedom From Bias of the Instructional Coaching Assessments

RS and ISS items were expected to function similarly across teacher demographic groups of age, educational degree, and years of teaching experience. Reynolds and Carson's (2005) partial correlation model for detection of DIF was used to identify potential bias in RS and ISS item scores based on the three teacher groups (i.e., age, educational degree, and years of teaching experience), controlling for Total RS scores. Overall, the RS items did not evidence DIF for the three teacher demographic groups. For teacher age, educational degree, and years of teaching experience, the RS of Quality Instruction, Positive Behavior Management, and Responsive Learning Communities item partial correlations were small, ranging from –.18 to .18, and not statistically significant for the Total and three subscales of Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills. Similarly, the ISS of PCC and NCC item partial correlations were small and nonsignificant, ranging from –.15 to .17. In sum, the RS and ISS items were found to function comparably across age, educational degree, and years of teaching experience.

Usability Evidence for the PMR

At the end of each PMR, coaches and teachers were asked to rate on a 5-point Likert-type scale (0 = *strongly disagree* to 4 = *strongly agree*) 10 usability items (see Table 3).

Overall, both users rated the PMR as easy to use with clear instructions (Q3 and Q9). In general, teachers viewed the PMR slightly more favorably than coaches, especially when asked about the positive consequences of using the PMR (Q6), whether they could complete the PMR during the coaching process (Q7), and the overall social validity (Q10). The total scores for coaches (*n* = 27) and teachers (*n* = 177) were 62.59 (*SD* = 11.51) and 68.02 (*SD* = 11.47), respectively. Normative interpretations on the acceptability of total scores indicate that total scores in the 60s are in the marginally acceptable range (Bangor et al., 2008).

Discussion

This article described a new measurement model for assessing instructional coaching, examined psychometric evidence for the Instructional Coaching Assessments–Teacher Forms, and illustrated feedback coaches can use to improve their effectiveness and interactions with teachers. The perspectives of teachers were featured in this initial investigation given that most instructional coaching is designed to facilitate continuous PD of teachers. The measurement framework defines and quantifies specific actions and outcomes of coaching effectiveness that are related to improved instruction and student achievement. As illustrated, the Instructional Coaching Assessments–Teacher Forms are designed to provide an evaluation of coaching effectiveness, interactions, and skill-specific strengths and areas for improvement.

The present investigation examined the internal structure of the assessments in the context of a theoretical framework and foundational reliability and validity inferences for scores from the RS and ISS–Teacher Forms. The current findings provided evidence to support the reliability of scores from the RS and ISS, as well as the validity of ensuing score inferences about coaching effectiveness and interpersonal interaction characteristics. Specifically, reliability and validity evidence suggested high internal consistency, supported the RS hypothesized three-factor model (see Table 1) and ISS hypothesized two-factor model (see Table 2), and relative freedom from item bias. In addition, coaches and teachers evaluated the utility of the PMR, an assessment embedded throughout the three-phase coaching process, as easy to use and helpful (see Table 3). Teachers further agreed that the use of the PMR can have positive consequences on the coaching experience and that they can be completed at three points during the coaching process.

Initial Psychometric Findings

Our current findings support the proposed dimensional structure of the RS and ISS–Teacher Forms, reflecting evidence for the RS hypothesized three-factor model and ISS hypothesized two-factor model. Specifically, CFA models

for the RS of Quality Instruction, Positive Behavior Management, and Responsive Learning Communities were found to be acceptable and support the use and interpretation of the RS Total, subscales, and item-level action scores. Although CFA findings suggest that the RS subscales of Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills are acceptable measures of teachers' perception of coaching effectiveness, the action scores may be more interpretable and useful for informing individual and school-wide PD planning. Similarly, results from principal components analyses of the ISS offered support for the two-factor structure—PCC and NCC—used to report prevalence of interpersonal aspects of the coaching process.

The precision of the scores from both the RS and ISS–Teacher Forms was good. Specifically, the internal consistency of both the RS and ISS were high as evidenced by Cronbach's alphas for the Total, subscales, and/or action scores, and the resulting SEMs were very low for virtually all scales and subscales. Although no measure on its own should be used for high-stakes decision making, all scores from the RS and ISS had Cronbach's alphas exceeding .95 typically considered acceptable for such decisions. Also, item-to-total correlations were strong (exceeding .70) for each scale (Obermiller & Spangenberg, 1998). In summary, the initial internal consistencies, low SEMs, and item-to-total scale relations are sufficient for the measure's intended purpose, to gather reliable feedback on the effectiveness of instructional coaching for informing PD from teachers.

Evidence of relations to other variables at the item level was positive in support of the RS and ISS. Item bias is an important consideration for measurement development, as it indicates whether membership in various groups is likely to systematically influence the relationship between item ratings and the underlying construct. The RS and ISS items were assessed for bias based on teachers' age, years of experience, and educational degree, as well as input from expert and consumers. Years of teaching experience and educational degree are commonly examined predictors of student achievement (Odden, 2008). Both scales did not evidence item bias based on teacher groups, suggesting that items functioned similarly regardless of age, teaching experience, or educational degree. These findings support that the RS and ISS scores reflect teacher perceptions of coaching effectiveness or prevalence, independent from demographic group influence, and support the intended inferences.

Implications of Coaching Formative Assessment for PD Improvement

Instructional coaches are leaders in delivering effective job-embedded PD that meets the complex needs of educators *and* students (Denton & Hasbrouck, 2009; Glover & Reddy, 2017; Reddy et al., 2017). Becoming an effective coach, however, constitutes high levels of proficiency in problem

solving, data use and interpretation, modeling, facilitative practice, performance feedback, and overall interaction style that, in combination, effectively and efficiently can result in professional growth for educators and school improvement. Thus, an effective coach not only requires specialized training but also ongoing support and accurate feedback that is useful, specific, and immediate (e.g., Hagermoser Sanetti et al., 2007; Noell et al., 1997; Reddy, Dudek, & Lekwa, 2017). For coaches to receive such support, it is necessary that relevant skill-based needs can be assessed efficiently and reliably over time to generate actionable feedback.

The findings from this study constituted a first step in developing and validating a multirater assessment for instructional coaching in schools. As mentioned, few technically sound formative assessments of instructional coaching exist (i.e., Lane et al., 2013), often resulting in coaches being evaluated using tools and standards meant for teachers. As Killion et al. (2012) stated, "When coach evaluations use teacher standards, however, principals or other supervisors must extrapolate to apply those standards to coaching work, potentially making evaluations inconsistent" (p. 136). This void in the available assessments impedes school leaders from accurately monitoring the performance of instructional coaches and effectively tailoring PD to advance coaches' competencies and to support teachers' efforts to improve services for their students. Thus, the potential impact of coaching formative assessments to inform coaching effectiveness, which in turn enhances educators' practices and student outcomes, is significant for school improvement.

iCoach assessments offer several unique features that may aid in coaching and teacher PD efforts. First, they measure specific coaching actions found to be effective through rigorous research at improving teacher and student outcomes. The framework driving these actions has been supported by decades of consultation and coaching research to increase (a) the fidelity and acceptability of implementation of research-based practices for improving classroom behavioral management and instructional interventions and (b) students' behavioral and academic performance (e.g., Fabiano et al., 2018; Glover, 2017; Reddy, Dudek, & Lekwa, 2017).

Second, iCoach defines and quantifies coaching effectiveness within an action–outcome measurement framework. This framework can help stakeholders (i.e., school leaders, coaches, and teachers) define what constitutes effective instructional coaching, develop a shared understanding of the coaching process, and delineate the key roles and responsibilities for this important leadership role in schools. This is an essential step in determining how instructional coaching situates in various school contexts and how the intended roles, activities, and outcomes of coaching align with school improvement plans.

Third, iCoach assessments are evidence-centered in that they require all informants to identify sources of evidence (i.e., opportunities to observe and interact with the coach across context) before evaluating coaching. The collection of sources of evidence is critical to ensuring evaluation is comprehensive, valid, and meaningful for informing professional improvement.

In addition, iCoach generates online comprehensive feedback reports for individual coaches and school districts within a year and across school years that enable school leaders to track professional improvement for coaches across or within grade levels and year. Reports present scores and visual performance feedback of overall effectiveness and development in skill-focused coaching actions, which can be used in concert with other data sources such as teacher effectiveness evaluations and student achievement, to inform decisions for school improvement plans. Moreover, this information promotes timely and accurate decision making regarding supports needed for improvements in coaching practices in school systems.

Limitations

Although advancing fundamental psychometric evidence for two iCoach assessments, this study was characterized by some common limitations. First, the sample size in the present investigation allowed us to focus only on teacher data and prevented the examination of the psychometric properties of the RS and ISS with other informants—coaches and supervisors of coaches—in our 360 assessment model. Second, independent observations of coaches' actions and interactions were not conducted, precluding a comparison of direct observational data to indirect behavior ratings from the perspective of teachers. Third, this initial investigation was limited to 11 high-poverty, low-performing charter schools in the State of New Jersey. The participating schools represented high rates of staff turnover and stress, as well as below benchmark performance on statewide student achievement tests. Thus, results from this investigation may be less generalizable to other school communities in the nation. Despite this limitation, teacher demographics (e.g., ethnicity, race, and tenure status) were comparable with estimates across New Jersey. Finally, many of the instructional coaches previously worked as teachers in the schools where they were subsequently coaching and thus had already established relationships with teachers. This might have influenced ratings.

Future Research

Research on the validity of coaching assessments offers new avenues of investigation. For example, studies are needed that examine the preferred RS and ISS factor models to competing alternative higher order models. Similarly,

the sample size in this study prevented the authors from testing higher order factor models for the three RS (i.e., Quality Instruction Total, three subscales, and six coaching actions), which are needed to fully validate the internal structure. Furthermore, there are a number of additional combinations of models that might be considered, and the conclusions on model fit are limited to those examined in this present study. Investigations that examine different methods for scoring the RS and ISS–Teacher Forms would be beneficial. Also, research that examines PMR ratings in relation to the end-of-year summative RS ratings is warranted. Future studies that examine sources of evidence used by raters to evaluate coaching effectiveness and interactions will offer validity evidence for content and response processes. In another study, the sensitivity of score change following coaching trainings and supports, or data-driven performance feedback should be evaluated. Also, investigations that examine cross-informant agreement of iCoach assessments are needed. Finally, predictive validity studies can be conducted that examine the ability of scale scores to predict teacher effectiveness and student achievement as measured on formative assessments and/or state testing.

Conclusion

Given attention in the ESSA (2015) to educator evaluation, job-embedded PD, and coaching, formative assessments that generate valid results for coaches are needed that clearly define and quantify the process, actions, and outcomes of effective coaching in schools. This investigation served as a first step in examining the use of assessments completed by teachers of their instructional coaches. Research is warranted that examines how an assessment-driven feedback system may aid in the development of coaching excellence, and ultimately school improvement. As instructional leaders, coaches play a key role in supporting teachers in enhancing the quality of instruction and behavior management that met the continuum of needs of educators *and* students in schools.

We synthesized the theoretical and empirical foundations for the development of a comprehensive coaching assessment. The initial psychometric findings for the RS and ISS–Teacher Forms were promising. School personnel interested in teacher evaluation and development of coaching talent now have assessments that can yield reliable scores that permit valid inferences about the extent to which their instructional coaches effectively implement research-based coaching actions. In addition, teachers can assess key characteristics of coaching interactions. Collectively, this information can be used to guide the PD of coaches and likely enhance their effectiveness at collaborating with teachers to implement data-based interventions. Additional, large-scale validations will be useful for further testing the internal structure, reliability, and external validity of this assessment system, as well as examining the efficacy of this assessment system on coach and educator effectiveness at improving student achievement.

ORCID iD

Todd Glover (iD) https://orcid.org/0000-0001-7100-8139

References

American Institutes for Research. (2005). *Conceptual overview: Coaching in the professional development impact study*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

APQC Education Advanced Working Group. (2011). *Best practices in evaluating the impact of instructional coaching on teaching and learning* (Final report). Retrieved from https://www.apqc.org/advisory-services/advanced-working-groups

Arbuckle, J. L. (2015). IBM SPSS Amos 21 user's guide (p. 635). Crawfordville, FL: Amos Development Corporation.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 6, 574–594.

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles, CA: BMDP Statistical Software.

Bergan, J. R., & Kratochwill, T. (1990). *Behavioral consultation and therapy*. New York, NY: Plenum Press.

Bergan, J. R. (1977). *Behavioral consultation*. Columbus, OH: Charles E. Merrill.

Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, 51, 5–32.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. A. Thomas, Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London, England: Taylor & Francis.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.

Comrey, A., & Lee, H. (1992). *A first course in factor analysis*. Hillsdale, NY: Lawrence Erlbaum.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart and Winston.

Denton, C. A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation*, *19*, 150–175.

Deussen, T., Coskie, T., Robinson, L., & Autio, E. (2007). *"Coach" can mean many things: Five categories of literacy coaches in reading first* (Issues & Answers, REL 2007-No. 005). Portland, OR: Regional Educational Laboratory Northwest.

Erchul, W. P. (2015). Put me in, coach: Observations on selected studies implementing supportive interventions to teachers. *School Mental Health*, *7*, 74–79.

Erchul, W. P., & Sheridan, S. (2014). *Handbook of research in school consultation*. New York, NY: Routledge.

Every Student Succeeds Act of 2015 (ESSA), 20 U.S.C § 6311. et seq. (2015).

Fabiano, G., Reddy, L. A., & Dudek, C. M. (2018). Use of teacher coaching and formative assessment for improving classroom practices. *School Psychology Quarterly*. Advance online publication. doi:10.1037/spq0000223

Finstad, K. (2006). The System Usability Scale and non-native English speakers. *Journal of Usability Studies*, *4*, 185–188.

Glover, T. A. (2017). A data-driven coaching model used to promote students' response to early reading intervention. *Theory Into Practice*, *56*, 13–20.

Glover, T. A., & Reddy, L. A. (2017). Instructional coaching: Consideration of key competencies from multiple models. *Theory Into Practice*, *56*, 1–2.

Gulamhussein, A. (2013). The core of professional development. *American School Board Journal*, *197*, 22–26.

Hagermoser Sanetti, L. M., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification*, *31*, 454–465.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55.

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Uppsula, Sweden: Scientific Software.

Killion, J., Harrison, C., Bryan, C., & Clifton, H. (2012). *Coaching matters*. Oxford, OH: Learning Forward.

Kurz, A., Reddy, L. A., & Glover, T. (2017). A multidisciplinary framework of instructional coaching. *Theory Into Practice*, *56*, 66–77.

Lane, M., Robbins, M., & Price, D. (2013). Establishing content validity for a literacy coach performance appraisal instrument. *Literacy Research and Instruction*, *52*, 230–253.

Mead, S. (2012). *Recent state action on teacher effectiveness: What's in state laws and regulation?* New York, NY: Bellwether Education Partners.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1), i–29.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, *12*, 77–88.

Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology*, *7*, 159–186.

Odden, A. (2008). *How to create world class teacher compensation*. Saint Paul, MN: Freeload Press.

Reddy, L. A., Barboza-Whitehead, S., Files, T., & Rubel, R. (2000). Clinical focus of consultation outcome research with children and adolescents. *Special Services in the Schools*, *16*, 1–22. doi:10.1300/J008v16n01_01

Reddy, L. A., Dudek, C., & Lekwa, A. J. (2017). Classroom strategies coaching model: Integration of formative assessment and instructional coaching. *Theory Into Practice*, *56*, 46–55.

Reddy, L. A., Dudek, C., & Shernoff, E. (2016). Teacher formative assessment: The missing link in response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.; pp. 607-623). New York, NY: Springer Science.

Reddy, L.A., Fabiano, G., Dudek, C., & Hsu, L. (2013). Development and construct validity of the Classroom Strategies Scale. Special Series: Assessment of general education teachers' tier 1 classroom practices: Current science and practice. *School Psychology Quarterly*, *28*, 317-341. doi:10.1037/spq0000043

Reddy, L. A., Glover, T. A., Elliott, S., & Kurz, A. (2016, July). *Measurement of instructional coaching effectiveness: Gaps in assessment practices and validity*. Paper session: Assessment for Learning and Teachers' Assessment Practices in the Classroom Contexts at the International Testing Commission, Vancouver, British Columbia, Canada.

Reddy, L. A., Kettler, R. J., & Kurz, A. (2015). School-wide educator evaluation for improving school capacity and student achievement in high-poverty schools: Year 1 of the school system improvement project. *Journal of Educational and Psychological Consultation*, *25*, 90–108.

Reddy, L. A., Shernoff, E. S., & Lekwa, A. L. (2017, February). *Examining the efficacy of the collaborative coaching model implemented in high poverty schools*. In Holdaway, A. (Chair), Coaching Teachers in Universal SEB Programming: Promising Research-Practice Models. Symposium submitted to the annual meeting of the National Association of School Psychologists, San Antonio, TX.

Reynolds, C. R., & Carson, D. (2005). Methods for assessing cultural bias in tests. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 795–823). New York, NY: John Wiley.

Sawchuk, S. (2016, January). ESSA loosens reins on teacher evaluations, qualifications. *Education Week*, *35*(15), 14–15.

Schwartz, F., & White, K. (2000). Making sense of it all: Giving and getting online course feedback. In K. W. White & B. H. Weight (Eds.), *The online teaching guide: A handbook of attitudes, strategies, and techniques for the virtual classroom* (pp. 57–72). Boston, MA: Allyn & Bacon.

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Garbacz, S. A., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of Conjoint Behavioral Consultation and the mediating role of the parent-teacher relationship. *School Psychology Review*, *41*, 23–46.

Sheridan, S. M., & Kratochwill, T. R. (2008). *Conjoint behavioral consultation: Promoting family-school connections and interventions*. New York, NY: Springer.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*, *45*(3), 12–16.

Shute, V. (2007). *Research report: Focus on formative feedback*. Princeton, NJ: Educational Testing Service.

Shute, V., Kim, Y. J., & Razzouk, R. (2013). *ECD for dummies*. Princeton, NJ: Educational Testing Service.