PILOT VERSION OF THE iCOACH ASSESSMENT SYSTEM: A PRELIMINARY INVESTIGATION OF PSYCHOMETRIC PROPERTIES

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

ALEXANDER ALPERIN

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY OCTOBER 2017

APPROVED: _____
Linda A. Reddy, Ph.D.

_____
Adam Lekwa, Ph.D.

DEAN: _____
Francine P. Conway, Ph.D.

Abstract

Research on assessment instruments designed to formatively evaluate instructional coaches is sparse. Twenty-six assessments that evaluate the performance of instructional coaches were found through two comprehensive searches (i.e., internet search and literature search) and only one of these assessments (i.e., Lane, Robbins, & Price, 2013) appears to have supporting psychometric evidence in a peer-reviewed publication. This study presents initial validity evidence of the iCoach Assessment System (Reddy, Glover, Elliott, & Kurz, 2016), an on-line, multi-rater, research-based tool for evaluating instructional coaches' competency level and implementation skill. Analyses of data from 114 participants (105 teachers and 9 instructional coaches) were conducted to examine the initial reliability and validity of the iCoach Assessment System scales (scales: Quality Instruction, Behavior Management, and Responsive Learning Communities). The iCoach Scales were hypothesized to have three common factors: (1) Goal Formulation Skills, (2) Implementation Support Skills, and (3) Evaluation Skills, which encompass six coaching actions (coaching actions: identifying needs and resources, setting goals, designing implementation plans, modeling implementation steps, providing performance feedback, and evaluating implementation and goal attainment). Results revealed that the pilot version of the iCoach Scales and hypothesized factors yielded acceptable internal consistency estimates, item-to-total correlations, and freedom from item bias with teacher demographics (i.e., age, year of experience, and degree). Principal components analyses (with varimax rotation) suggested that each iCoach Scale yielded multiple factors. Implications for research and practice are discussed.

Acknowledgements

I would like to express gratitude to a number of people for their support throughout my doctoral studies and the dissertation process. First and foremost, I would like to thank my dissertation chair, Dr. Linda Reddy, who has truly been a mentor as well as a constant source of support. Her work ethic was inspiring and I am grateful to have had her play such an influential role in my graduate education. I could not have asked for a better dissertation chair. I would also like to thank Dr. Adam Lekwa, who has been very supportive and helpful with my dissertation. I am grateful that he was always able to answer phone calls and e-mails in order to assist with my dissertation.

I would like to thank my mother and stepfather, Linda Lande and Richard Brown; my father and stepmother, Dr. Richard Alperin and Barbara Holman; and my sister, brother in-law, and nephews, Dr. Heather Bohn, Adam Bohn, Ezra Bohn, and Ezekiel Bohn. I appreciate and am fortunate to have had their guidance and support. Their constant encouragement helped me more than they know.

Thank you to Dr. Evelyn Liegner. I appreciate her support and help throughout the years. Additionally, I am grateful to my best friends- Andrew Howard, Buddy Marsico, Chris Deflora, Alex Maro, Jordan Young, and Damian Fumero. I am truly blessed to have such amazing friends and thank you for the long phone calls, jokes, and support. I love you all.

Table of Contents

LIST OF TABLES

Introduction

National school reform and educator evaluation have increasingly emphasized teacher effectiveness and accountability in order to improve achievement for all students (Duncan, 2012). Similarly, federal grant programs such as Race to the Top (2009; 2010), the Teacher Incentive Fund (2010; 2012) and amendments to the Elementary and Secondary Education Act (ESEA; U.S. Department of Education, 2012) or No Child Left Behind (NCLB) waivers have required that teacher evaluation become more rigorous, comprehensive, and based on multiple measures (Reddy, Dudek, Kettler, Kurz, & Peters, 2016). Recently, teacher evaluation practices were further impacted by the authorization (December, 2015) of the Every Student Succeeds Act (ESSA), which has placed responsibility to state education departments for teacher evaluation. In response to the recent attention on school and teacher evaluation reform and the legislation of the ESSA, many states in the United States have and will continue to implement significant policy changes in an attempt to improve teacher evaluation practices (Mead, 2012; Sawchuk, 2016). The rationale behind these changes is that teacher evaluation reform will lead to improved classroom practice (e.g., high quality instruction), which in turn, will result in better outcomes for student achievement.

To strengthen teacher evaluation and classroom practices, professional development (PD) has been highlighted as a crucial element for teacher success (Beglau et al., 2011). This is demonstrated by the 2001 NCLB Act, which requires school districts to institute PD programs for schools that fail to make sufficient yearly progress for two years or more (Kowal & Steiner, 2007). This emphasis on PD is not only apparent in legislation, but also in the significant expenditures that school districts devote to PD. Resnick (2010) cited in a recent survey of PD trends that the average teacher in the United States received 25.4 hours of PD annually and the

average school district invested $225, 200 annually in PD. Projecting the results of this survey out to the estimated 16,000 school districts nationwide, PD is likely at minimum a $3.6 billion dollar industry (Beglau et al., 2011).

Considering the substantial resources invested in PD for teachers, it is imperative that effective and efficient methods be used to promote teacher competencies and implementation in the classroom. In contrast to traditional full-day one time workshops, school personnel and researchers are seeking job-embedded, individualized approaches that yield sustainable and informed site-based and content-specific expertise (e.g., Denton & Hasbrouck, 2009; The Commission on Effective Teachers and Teaching [CETT], 2012; Wren & Vallejo, 2009). Instructional coaching has emerged as one PD strategy that fits this criterion. Instructional coaching typically consists of classroom teachers and curriculum specialists assuming leadership roles in their schools and training and supporting their colleagues (Denton & Hasbrouck, 2009). While many instructional coaches assist with curriculum planning, the current research is consistent with Shernoff, Lakind, Frazier, and Jakobsons' (2015) operationalization of instructional coaching as "job-embedded, sustained, classroom-based support" that enhances teachers' "instructional skills and use of evidence-based practices" (p. 2). Instructional coaching usually includes some form of teacher observation (Denton & Hasbrouck, 2009). It is important to note that there is a wide range of instructional coaching models used in schools (e.g., Denton & Hasbrouck, 2009; Rush & Young, 2011; Stormont, Reinke, Newcomer, Marchese, & Lewis, 2015). Likewise, there is a lot of variability in the titles and responsibilities for instructional coaches. For example, the term instructional coach has been used synonymously with instructional facilitator, instructional specialist, peer coach, peer assistant, mentor teacher, literacy coach, mathematics coach, etc.

While the variability in definitions and applications has presented a challenge to researchers, emerging evidence in support of instructional coaching does exist in the education literature. During the past decade, research has focused on reading and literacy coaching with a handful of large-scale randomized control trials. For example, Powell, Diamond, Burchinal, and Koehler (2010) found positive instructional and student achievement outcomes for two forms of early literacy coaching. Similarly, Vernon-Feagans, Kainz, Hedrick, Ginsberg, and Amemdum (2013) reported that struggling readers in schools offering coaching significantly outperformed struggling readers in control schools. Researchers have also begun to find evidence that instructional coaching is effective in mathematics (e.g., Campbell & Malkus, 2011), science (e.g., Lee, DeChenne, Nugent, Kunz, & Houston, 2014), behavioral management (e.g., Hershfeldt, Pell, Sechrest, Pas, & Bradshaw, 2012), and teacher-child interactions (Pianta, Mashburn, Downer, Hamre, & Justice, 2008). Despite the emergence of evidence for instructional coaching in the past ten years, more research is needed that both directly supports the effects of coaching on teachers' classroom practices *and* student academic achievement (Cornett & Knight, 2009; Denton & Hasbrouck, 2009; & Stormont, Reinke, Newcomer, Marchese, & Lewis, 2015). Even though more research is needed regarding instructional coaching, there is an abundance of evidence at an instructional coach's disposal such as those regarding effective teacher and classroom practices and there is continually accumulating evidence concerning strategies and methods to positively influence a teacher's professional development (Briere, Simonsen, Sugai, & Myers, D, 2015; Sutherland & Wehby, 2001).

The emphasis on instructional coaching is recognized at the federal level. For example, the National Center for Special Education Research (NCSER) explicitly states that they "are still working to understand the key elements of effective coaching" (U.S. Department of Education,

Institute of Education Sciences [IES], 2014, p. 4). In addition, the IES (2016) is currently offering education research grants for studies related to the topic of Effective Teachers and Effective Teaching, which is concerned with exploring strategies (e.g., instructional coaching) for improving the performance of the classroom teacher.

While there is a lack of agreement on titles and roles for instructional coaching, coaching is growing at a substantial and staggering rate in schools (Knight, 2006; Wren & Vallejo, 2009). The Center on Education Policy (2006) found that for 2004-2005, 60 percent of districts had used instructional coaches (referred to as "distinguished teachers") to assist struggling schools. The International Reading Association et al. (2006) reported that "coaching has been adopted as the model for professional development in Boston, Dallas, New York, and Philadelphia public schools" (p. 35). The National Education Association (NEA; 2011) asserted that instructional coaching (referred to as "Peer Assistance") is being increasingly used as a PD strategy. United States Secretary of Education, Arne Duncan, endorsed instructional coaching through his comments on Peer Assistance Review (an approach that appoints expert or consulting teachers who mentor and support other teachers), which was described by him as "where the country needs to go" (Duncan, 2009; Papay & Johnson, 2012; Winerip, 2011). Likewise, President Obama and the American Federation of Teachers President Randi Weingarten have both deemed Peer Assistance Review to be an approach with great potential for improving professional evaluation and teacher quality (e.g., Dillon, 2008; Goldstein, 2004, 2007; Koppich, 2005; Obama, 2009; Papay & Johnson, 2012; Toch & Rothman, 2008).

As instructional coaching has emerged as a popular model for ongoing, job-embedded PD to teachers; having adequately trained and effective instructional coaches is critical for promoting teacher effectiveness and student learning (Lane, Robbins, & Price, 2013). Effective

coaches present proficiencies in problem solving, data use and interpretation, performance feedback, and overall interaction style that in combination effectively and efficiently result in professional growth for educators. Thus, becoming an effective coach not only requires specialized training, but also on going performance assessment and data-based feedback which is useful, specific, and immediate (e.g., Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Sanetti, Luiselli, & Handler 2007). An evidence-based assessment that targets coaches' skills and performance over time would offer a method for identifying areas of strength and areas needing improvement (Lane et al., 2013). Despite the need for coach assessments, few currently exist. This marks a significant gap in the availability of school personnel evaluation tools. Below is a brief review of the availability of assessments for instructional coaches.

Literature Search Approaches

Two comprehensive search approaches were conducted on assessments for instructional coaching in education: a literature search from 1984 through the present and an internet search of publically available assessment tools. For both search approaches, assessment instruments for teachers mentoring student teachers and coaches guiding principals were not included; rather the focus was solely on assessment instruments that pertained to instructional coaches who mentor teachers in schools.

For the literature search (i.e., peer reviewed journals and book chapters), the following key words were used: "instructional coach," "instructional specialist," "peer coach," "peer assistance," "mentor teacher," "education coach," "literacy coach," "survey," "instrument," "evaluation," and "rubric." The databases used were Articles+ (which includes hundreds of databases such as Academic Search Premier, JSTOR, Science Direct, and Web of Science), Google Scholar, and ProQuest. In addition, a review of selected peer-reviewed journals known to

publish literature pertinent to the fields of education and psychology was conducted to ensure a comprehensive literature search (e.g., *American Education Research Journal, American Journal of Educational Studies, School Psychology Quarterly, School Psychology Review, Journal of School Psychology, Journal of Education Psychology, and Review of Educational Research).* Following the initial search, eight articles were found. After closer examination of the articles (e.g., reading the abstract), it was found that seven of these articles did not meet the inclusion/exclusion criteria described above. In particular, three of the articles pertained to business/executive coaching, one article concentrated on coaches in healthcare, one article focused on athletic coaching, one article dealt with student teachers, and one article attended to coaches mentoring principals.

The literature search resulted in one study (See Table 1 for additional detail on the study) from Lane et al. (2013) that reported on the development of two Literacy Coach Appraisal Instruments- a "Long Form" and a "Short Form"- as a result of a factor analysis reflecting evidence of content validity (Lane et al., 2013, p. 242).

The internet search was conducted using *Google* with the following key terms: "instructional coach," "instructional coach evaluation," "instructional coach evaluation form," "instructional coach evaluation rubric," "instructional coach evaluation tool," "instructional specialist," "instructional specialist evaluation," "instructional specialist evaluation form," "education coach evaluation instrument," "mentor teacher evaluation rubric," "mentor teacher evaluation instrument," "peer coach evaluation rubric," "peer coach evaluation instrument," "literacy coach evaluation instrument," and "literacy coach evaluation rubric."

This search resulted in a total of 24 publically available assessment instruments (See Table 1 for additional details). These instruments were found on a wide variety of websites,

including: state departments of education (i.e., Pennsylvania Department of Education and Arkansas Department of Education), school districts, public schools, charter schools, nonprofit educational organizations, and practitioners/researchers (e.g., James Knight). Five of these instruments were specifically designed to evaluate Literacy Coaches. Five of the instruments were self-assessments. Two of the assessment instruments were found on the *Examining Mathematics Coaching (EMC) Project* website (i.e., Yopp, Burrough, & Sutton, 2010a; Yopp, Burroughs, & Sutton, 2010b), which is a five-year research study (2009-2014) that investigated effective coaching in K-8 mathematics classrooms. Another assessment instrument was identified from the *Florida's Problem Solving/Response to Intervention (PS/RtI) Project* (i.e., Castillo et al., 2013), which is an ongoing collaboration between the Florida Department of Education and the University of South Florida to support and evaluate the implementation of the PS/RtI model in a select group of schools. While the instruments and the corresponding psychometric data from the *EMC Project* and the *PS/RtI Project* were presented on each of their websites, neither was published in a peer-reviewed publication.

A majority of the instruments that were found resemble Danielson's (2012) *Framework for Instructional Specialist Evaluation Instrument*. The Danielson (2012) model evaluates the instructional coach (referred to as "Instructional Specialist") across four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. Each domain is comprised of specific competencies (19 in total across 4 domains), where the evaluator rates the instructional coach on a 4-point scale. Overall, reliability and validity evidence were found for only 3 of the 24 assessment tools (i.e., *EMC* and *Florida PS/RtI* tools).

Taken as a whole, research on assessment instruments designed to be used to formatively evaluate instructional coaches is sparse. Few assessments exist that were designed to evaluate the

performance of instructional coaches and even fewer of these assessments are shown to be reliable and valid (i.e., Lane et al., 2013). Presently, schools are using invalid or modified assessments for instructional coaches. Even though international and national leadership and content area standards exist for instructional coaches, these standards are nearly ten years old and are not used in any coach evaluation framework or assessment (i.e., International Reading Association et al., 2006). Due to the gap in available instruments in coach assessments, instructional coaches are often evaluated using standards meant for teachers. As Killion, Harris, Bryan, and Clifton (2012) state, "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). The lack of reliable and valid assessments, specifically designed to evaluate the job performance of instructional coaches, impedes school leaders from accurately assessing the performance of instructional coaches and effectively tailoring supports and professional development opportunities to advance coaches' skill and competencies (Lane et al., 2013). A measurement framework for evaluating instructional coaching is sorely needed to ensure that the coaching of educators is effective in meeting school, educator and student achievement needs. This study presents the development and initial evidence of the pilot version of the iCoach Assessment System (Reddy, Glover, Kurz & Elliott, 2015), an on-line, multi-rater, research-based tool for evaluating instructional coaches' competency level and implementation skills.

Current Study

The present study was conducted to examine the preliminary evidence of the internal structure of the iCoach Assessment System (Reddy, Glover, et al., 2015). The study had three aims: (a) describe validity evidence based on internal structure and construct validity; (b)

describe validity evidence based on relations to other variables (i.e., item bias); and (c) compare the teachers' ratings of the instructional coaches to the instructional coaches self-ratings. It was hypothesized that internal consistency estimates would be in the acceptable for the iCoach Scales, factors, and cells (Cicchetti, 1994) and include acceptable item-to-total correlations (Obermiller & Spangenberg, 1998). It was hypothesized that the iCoach Scales, Quality Instruction, Behavior Management, and Responsive Learning Communities would each have multiple factors as evidenced through principal components analyses. It was also hypothesized that the instructional coaches would rate themselves more favorably compared to the teachers' ratings of them.

Methodology

Participants

The sample included 114 participants (105 teachers and 9 instructional coaches) from five charter schools within New Jersey that were surveyed during the spring of 2015 (See Table 2). From the aggregate sample, distinctive samples were identified based on participant type: an instructional coach sample (*n*=9) and a Pre-K through 8th grade classroom teacher sample (*n*=105).

Teacher sample. As shown in Table 2, there were 16 male (15.20%) and 89 female (84.80%) teachers that were coached and completed the iCoach assessments. The average age of teacher respondents was 32.21 years (SD= 8.69 years). The mean years of teaching experience was 5.40 (SD=5.84 years, range = 0 to 39 years). The majority of the sample (77.14%) reported being Caucasian, a smaller subsample (15.24%) identified as African American, and an even smaller subsample reported being Asian (3.81%) or classified themselves as "Other" (3.81%). The vast majority of the sample reported their ethnicity as non-Latino (94.30%), while a minority identified as Latino (5.70%). Teacher response rate was 97.22% (105/108).

Instructional coach sample. The average age of instructional coaches was 36.11 (SD=8.74). The instructional coach sample was entirely female (100%). Five of the instructional coaches reported being Caucasian (55.55%), while two reported being African American (22.22%), one reported being Asian (11.11%), and one reported being Asian and Caucasian (11.11%). Each of the nine instructional coaches was assigned a minimum of 12 teachers to coach throughout the school year. Based on the teachers who completed the iCoach (See Table 3), the breakdown is as follows: one instructional coach had 20 teacher respondents (19.00%); one had 16 teacher respondents (15.20%); one had 15 (14.30%); one had 13 (12.40%); two had

11 (10.50%); one had 9 (8.60%); one had 6 (5.70%); and one had 4 (3.80%). Instructional coach response rate was 100% (9/9).

Measure

The iCoach Assessment System is a measurement framework for evaluating instructional coaches (Reddy, Glover, Elliott, & Kurz, 2016). It is an on-line, multi-rater, research-based tool designed to evaluate instructional coach effectiveness and generate scores for developing evidence-based coaching actions and skills through continuous performance feedback. The iCoach was designed to be content neutral, assessing core coaching actions and processes found across common areas of expertise such as literacy, mathematics, and science.

As a multi-rater assessment, the iCoach includes three forms for the following: teachers who have been coached; the supervisor for the coach; and a coach self-assessment. The collective perspectives of these raters provide what is commonly referred to as a 360 evaluation because individuals "around the coach" (i.e., administrators and teachers) and including the coach have input into the evaluation (copyrighted 2016 by iCoach Learning Innovations) .

The pilot version of the iCoach asks informants to rate the competency of the coach on a 4-point Likert type item scale from (1) novice, (2) developing, (3) competent, to (4) accomplished. The pilot version includes 140 items that represent actions and outcomes (scales) for the instructional coach. The three outcomes (scales) are: (a) Quality Instruction (QI); (b) Behavior Management (BM); and (c) Responsive Learning Communities (RLC). There is also a Total Scale, which offers a global assessment of coaching competency aligning to four levels (i.e., Novice, Developing, Competent, and Accomplished).

The QI Scale (53 items) refers to the data-based instructional decisions and practices that maximize students' academic performance. There is much empirical support for quality

instruction as a coaching outcome. Marsh, McCombs, and Martorell (2012) found that instructional coaches were viewed as effective if teachers and principals perceived them to have influence on teachers' instruction. Furthermore, Sutherland and Wehby (2001) demonstrated that students will improve their academic skills if they are actively engaged and provided with frequent opportunities to respond to academic tasks. This emphasis on quality instruction is echoed by Oliver and Reschly (2007), who state that there is substantial evidence that students are more successful in school if they are taught in a manner that is responsive to their readiness levels (e.g., Vygotsky, 1986), interests (e.g., Csikszentmihalyi, 1997) and learning profiles (e.g., Sternberg, Torff, & Grigorenko, 1998).

The BM Scale (49 items) corresponds to data-based behavior management decisions and practices that improve social interactions and healthy participation in the classroom and school. Behavior Management is imperative since teachers' ability to manage the behavior of their students is critical for student achievement (Oliver & Reschly, 2007). Specifically, to manage student behavior, it is more important for teachers to emphasize desirable behaviors to their students rather than penalize undesirable behaviors (Emmer & Stough, 2001). Teachers who have problems with behavior management and discipline in the classroom are not effective and often report high levels of stress and symptoms of burnout (Berliner, 1986; Browers & Tomic, 2000; Espin & Yell, 1994). Taken altogether, it is essential that instructional coaches assist teachers with managing the behavior of their students.

The RLC Scale (38 items) is concerned with the identification and systematic sharing of information and resources that address needs and goals of communities of teachers and students. Nelson (2009) conducted a 5-year case analysis which demonstrated that teachers collaborating with each other resulted in instructional improvements. Similarly, Andrews and Lewis (2007)

found that a professional learning community (PLC) amongst teachers resulted in increased knowledge and had a significant and beneficial impact in the classroom. It is vital that instructional coaches work with teachers to facilitate a school culture that promotes collaboration, professional development, and data-based decision making.

Each scale comprises three coaching skill clusters (hypothesized factors), which are Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills. The Goal Formulation Skills contain two coaching actions: (1) Identifying Needs and Resources and (2) Setting Goals. The coaching action of Identifying Needs and Resources refers to the instructional coach communicating and working collaboratively with teachers to gather information to identify needs and resources that improves student performance and teacher practices. This information guides the formulation of goals for the three scales (i.e., QI, BM, and RLC). Setting Goals involves the instructional coach collaborating with teachers to identify and write measureable goals based on student functioning and teacher practice relative to performance expectations. The goals, which are essential for implementation, are measureable, specific, and reflect high expectations. The Implementation Support Skills are comprised of the actions of (3) Designing Implementation Plan and (4) Modeling Implementation Steps. Designing Implementation Plans concerns the coach collaborating with teachers to create specific steps needed to achieve goals. When designing implementation plans, the instructional coach considers teacher skills, possible resources, and potential barriers to ensure that the plans are successfully implemented. Modeling Implementation Steps has the instructional coach demonstrate steps specified in implementation plans for the teacher. The instructional coach then supports the teacher's practice of these steps. Finally, the Evaluation Skills are made up of the coaching actions of: (5) Providing Performance Feedback and (6) Evaluating Implementation and Goal Attainment. Providing Performance

Feedback involves the instructional coach using data to provide specific, positive, and timely feedback about plan implementation and goal attainment. Evaluating Plan Implementation and Goal Attainment refers to the instructional coach using data to make judgments about the integrity of plan implementation and the teacher's progress toward goal attainment. This information is used to continuously improve plans and assess goal attainment and coaching effectiveness.

The iCoach framework described generates total, scale, factor, and cell scores (Reddy, Glover, Elliott, & Kurz, 2016). The Total Scale Score is a global or overall rating of the coach's competency level and is comprised of three factor scores, which are Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills. Goal Formulation Skills (49 items) include: QI-Goal Formulation Skills (18 items; e.g., *"Observing student learning."*); BM-Goal Formulation Skills (16 items; *"Creating goals for teacher-student interactions."*); and RLC-Goal Formulation Skills (e.g., 15 items; *"Listening to needs of learning communities."*). Implementation Support Skills (40 items) contains the following: QI-Implementation Support Skills (13 items; *"Demonstrating instructional practices."*); BM-Implementation Support Skills (14 items; *"Developing strategies to maximize positive student"*); and RLC-Implementation Support Skills (13 items; *"Creating plans for schoolwide behavior management."*). Evaluation Skills (51 items) is made up of QI-Evaluation Skills (e.g., 22 items; *"Determining whether students met academic expectations."*); BM-Evaluation Skills (e.g., 19 items; *"Determining goal attainment for behavior management."*); and RLC-Evaluation Skills (e.g., 10 items; *"Evaluating how well learning communities are functioning."*). The three factor scores are made up of cell scores, which correspond to each action nested in proposed scales (e.g., QI: Identifying Needs and Resources; PBM: Providing Performance Feedback, etc.). To summarize, the iCoach yields a

Total scale score, scale scores (i.e., coaching outcomes), factor scores (i.e., coaching skills), and cell scores (i.e., coaching actions).

The iCoach was designed as an evidence-centered assessment (Mislevy, Almond, & Lukas, 2003). Evidence-centered assessment designs can yield both formative and summative assessments, aggregate information from various sources, and accurately estimate complex competencies and dynamic performances (Shute, Kim, & Razzouk, 2013). Evidence-centered assessment design consists of three models, which are: (1) the competency model; (2) the evidence model, and (3) the task model (Shute, Kim, & Razzouk, 2013). The competency model refers to the collection of knowledge, skills, and attributes that comprise a highly effective coach. The evidence model bridges the competency and task model in that it defines the evidence that is needed to support the competency model. The task model describes the tasks or actions that are able to elicit the evidence described in the evidence model. As previously described, the iCoach defines a framework for what constitutes a highly effective coach. In evidence-centered assessment design, this framework is the competency model. Within this framework, the iCoach specifies which actions (e.g., identifying needs and resources), tasks, or situations a highly effective coach engages in, which can be thought of as the task model. Finally, ratings about these actions, tasks, or situations provide evidence for the claim that a certain coach is ineffective, partially effective, effective, or highly effective. On the iCoach, this is accomplished through the coach, teacher, and supervisor rating the coach on a Likert scale across 140 items.

The iCoach has raters indicate the type of evidence used to evaluate the coaches for all items. Specifically, after completing the 140 items, the rater is presented with six possible response options (the rater can choose multiple options) and has to select the type of evidence he or she used to inform the ratings of the coach in each of the outcomes of QI, BM, and RLC.

These six possible response options include: (1) Report from Others, (2) Personal Observations, (3) School Documents, (4) Meetings, (5) Other Sources, and (6) No Evidence. Next, the rater is asked to indicate the most important source of evidence (i.e., if rater selected multiple response options in the previous question, he or she would indicate which was the most influential) for the rating of the coach in each outcome.

The construct and item development of the iCoach was guided by modern test theory and established psychometric standards (American Psychological Association [APA], American Educational Research Association [AERA], & National Council on Measurement in Education [NCME], 2014). Development methods included several methods for factor and item formation. Initial development of the iCoach factors and items was guided by three methods: (a) comprehensive review of scholarly, peer-reviewed publications in the fields of education, sports psychology, and business, (b) critique of other related scales, and (c) review of constructs and items by educational assessment and instructional scholars.

Procedure

The sample was part of a school reform grant funded by the U.S. Department of Education's Teacher Incentive Fund grant, School System Improvement Project awarded to Rutgers University (PI Linda A. Reddy, Ph.D.). Informed consent was obtained from all participants in accordance with the Institutional Review Board at Rutgers University. The survey was distributed online, through Qualtrics, an online software tool, in a self-administered questionnaire format with forced choice items. Participants were provides a small stipend for their participation.

Data Analytic Plan

Several data analytic methods were used in this study. First, internal consistency via Cronbach's alpha was calculated for each of the three iCoach Scale associated factors, and cell scores. Second, item-to-total correlation was also computed to assess each item's relation to the proposed construct. Using Obermiller and Spangenberg's (1998) standards, items with item-to-total correlations below .50 suggest possible items for removal from a scale. Item-to-total correlations were computed to examine relationships between each item and the total scale scores as well as each item with the three factors. Items assigned to specific scales were hypothesized to correlate more highly with that scale than other scales. For example, an item assigned to the QI scale should have a higher correlation with QI than it does to the scales of BM and RLC.

Third, principal components analysis with varimax rotation and Kaiser normalization was used to determine how many factors each of the iCoach Scales. The varimax method with Kaiser normalization rotation is the most common method for the orthogonal rotation, which posits that the factors are not correlated (i.e., variables displaying high loading values in one factor will then have smaller loading valued in regards to the other factors; Utley, 2011). Statistical procedures using SPSS 23.0 were applied to the data to evaluate the hypothesized factor structure of each scale.

Reynolds and Carson's (2005) partial correlation model for detection of differential item functioning (DIF) was used to identify potential bias in iCoach mean item scores based on teacher age, educational level, and years of teaching experience by partialling out Total Scale scores. According to Reynolds, Wilson, and Chatman (1984), "the partial correlation method…is the current method of choice in examining item bias" (p. 220). Finally, informant ratings (self-

report vs. teacher) were compared through descriptive statistics and two-sample t-tests with Bonferroni adjustment.

Results

Internal Consistency

The iCoach Total scale score yielded strong internal consistency (α=.997). Likewise, strong internal consistency estimates were found for the QI Scale (α=.992), BM (α=.994), and RLC Scales (α=.994). Within the QI Scale, the three coaching outcomes of Goal Formulation Skills (α=.978), Implementation Support Skills (α=.972), and Evaluation Skills (α=.984) yielded strong internal consistencies. Within the BM Scale, the internal consistencies were also strong: Goal Formulation Skills (α=.982); Implementation Support Skills (α=.977); and Evaluation Skills (α=.985). Finally, in RLC, Goal Formulations (α=.982), Implementation Support Skills (α=.987), and Evaluation Skills (α=.975) yielded strong internal consistencies.

Item-to-total correlation was computed for the Total scale scores for QI, BM, and RLC, and each item within the scales. For QI, the item-to-total correlations ranged from .77 to .91. For BM and RLC, the item-to-total correlations ranged from .78 to .94 and .80 to .95. For each of the three scales, the item-to-total correlations were acceptable based on Obermiller and Spangenberg's (1998) standards, which recommend removal of items that have item-to-total correlations that are below .50.

Factor Analysis

In exploring the construct validity of the Pilot version of iCoach Scales, principal components analyses (PCAs) were computed (Meyers, Gamst, & Guarino, 2013). PCAs were computed to examine the underlying dimensionality of the iCoach QI, BM, and RLC scales. PCA is data-driven and identifies items that do not measure an intended factor or that measure multiple factors at once, and thus, diminishes the psychometric properties of the iCoach. Due to a small sample size to item ratio and linear dependency among variables, a principal components

analysis (PCA) and varimax rotation with Kaiser normalization were conducted separately for QI 53 items, BM 49 items, and RLC 38 items.

Communalities are values that represent the amount of variation for each variable that is accounted for by the factors (Zillmer and Vuz, 1995). The range of values for communalities is 0.0 to 1.0, with 0.0 suggesting that common factors do not explain the variance of a particular variable, and 1.0 indicating that all of the variance of a particular variable is explained by the common factors (Cook, 2014). Each of the extracted communalities in each of the analyses were above .6, which suggests that each item shares much common variance with the other items.

Since it was hypothesized that each iCoach Scale would have a nested 3 factor Model (i.e., Goal Formulation Skills, Implementation Support Skills, and Evaluation Skills), the number of factors were restricted to three during varimax rotation. For QI, the eigenvalues showed that the three factors accounted for 77.75% of the total variance. The first factor explained 28.90% of the variance, the second factor explained 24.94% of the variance, and the third factor explained 23.90% of the variance. In regards to BM, the eigenvalues showed that the three factors accounted for 83.67% of the total variance. The first factor explained 31.20% of the variance, the second factor explained 26.64% of the variance, and the third factor explained 25.83% of the variance. Lastly, for RLC, the eigenvalues demonstrated that the three factors accounted for 87.33% of the total variance. The first factor explained 32.65% of the variance, the second factor explained 29.80% of the variance, and the third factor explained 24.88% of the variance.

Comrey and Lee (1992) offered an interpretation guideline for loadings in a rotated component matrix, which is the following: 1.00 to .710 is excellent; .709 to .630 is very good; .629 to .550 is good; .549 to .450 is fair, and .320 or lower is poor. Using a cutoff of .550, 47 of the 53 items loaded on one factor for QI (See Table 4). Of those 47 items, 22 of the items had

loadings that were excellent, 14 items were very good, and 11 items were good. Two items loaded on two factors. For the 49 items of BM, 41 of the items loaded on one factor (See Table 5). Of those 41 items, 8 of the items had loadings that were excellent, 20 of items were very good, and 13 items were good. Eight of the items loaded on two factors. Of the 38 items of RLC, 29 of the items loaded on one factors (See Table 6). For those 29 items, 14 of the items had loadings that were excellent, 7 of the items were very good, and 8 items were good. Eight items loaded on two factors.

Examining QI's factor loadings through the three factors reveals that 17 of the 18 items from QI-Goal Formulation Skills had their highest factor loadings on component 2; 11 of the 13 items from QI-Implementation Support Skills had their highest factor loadings on component 3; and 19 of the 22 items from QI-Evaluation Skills had their highest factor loadings on component 1. Analyzing BM's factor loadings in regards to the three factors reveals that 11 of the 16 items from BM-Goal Formulation Skills had their highest factor loadings on component 2; 9 of the 14 items from BM-Implementation Support Skills had their highest factor loadings on component 3; and 13 of the 19 items from BM-Evaluation Skills had their highest factor loadings on component 1. Inspecting RLC's factor loadings through the three factors reveals that 11 of the 15 items from RLC-Goal Formulation Skills had their highest factor loadings on component 1; 7 of the 13 items from RLC-Implementation Support Skills had their highest factor loadings on component 1; and all 10 items from RLC-Evaluation Skills had their highest factor loadings on component 2.

Freedom from Item Bias

Items from iCoach Scales of QI, BM, and RLC were expected to function similarly across teacher demographic groups. Reynolds and Carson's (2005) partial correlation model for

detection of differential item functioning (DIF) was used to examine potential bias in each scale's mean items in relation to the three teacher groups (i.e., age, educational degree, and years of teaching experience), while partialing out QI, BM, or RLC Total scale scores.

For teacher age, educational degree, and years of teaching experience, the QI item partial correlations were minimal (-.318 to .344) and not statistically significant after correction. Likewise, the BM item partial correlations were minimal (-.302 to .267) and not statistically significant after correction. The RLC item partial correlations were also minimal (-.301 to .291) and not statistically significant after correction. Based on item partial correlations for DIF, items from QI, BM, and RLC were found to function comparably across age, degree, and years of teaching experience.

Comparison Scores of Informant Ratings

As shown in Table 3, descriptive statistics were computed between teacher and coach self-report ratings by scale, factor, and cell score. In general, results revealed that teachers rated their coaches slightly more favorably than coaches. An independent-samples t-test was conducted to determine if there were differences between teachers and coaches for the following: Total Scale, each scale total (QI, BM, and RLC), and the factors within each scale (QI-Goal Formulation Skills, QI-Implementation Support Skills, QI-Evaluation Skills, BM-Goal Formulation Skills, BM-Implementation Support Skills, BM-Evaluation Skills, RLC-Goal Formulation Skills, RLC-Implementation Support Skills, and RLC-Evaluation Skills). There was homogeneity of variances, as assessed by Levene's test for equality of variances, for the Total Scale (p=.134), QI (p=.069), RLC (p=.793), QI-Implementation Support Skills (p=.159), QI-Evaluation Skills (p=.132), BM-Evaluation Skills (p=.284), RLC-Goal Formulation Skills (p=.614), RLC-Implementation Support Skills (p=.824), and RLC-Evaluation Skills (p=.114).

The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances, for BM (p=.039), QI-Goal Formulation Skills (p=.038), BM-Goal Formulation Skills (p=.004), and BM-Implementation Support Skills (p=.015).

For the Total Scale, comparison of scores on the iCoach for teachers (M=3.493, SD=0.498) and coaches (M=3.284, SD=0.337) revealed no significant differences between groups, t(112)=1.235, p=.219. For QI, comparison of scores on the iCoach for teachers (M=3.504, SD=0.504) and coaches (M=3.260, SD=0.322) revealed no significant differences between groups, t(112)=1.423, p=.158. For BM, comparison of scores on the iCoach for teachers (M=3.496, SD=0.515) and coaches (M=3.390, SD=0.304) revealed no significant differences between groups, t(12.354)=.935, p=.368. For RLC, comparison of scores on the iCoach for teachers (M=3.449, SD=0.551) and coaches (M=3.084, SD=0.563) revealed no significant differences between groups, t(108)=1.898, p=.060.

For QI-Goal Formulation Skills, comparison of scores on the iCoach for teachers (M=3.533, SD=0.500) and coaches (M=3.233, SD=0.304) revealed significant differences between groups, t(12.080)=2.666, p=.020. For QI-Implementation Support Skills, comparison of scores on the iCoach for teachers (M=3.432, SD=0.577) and coaches (M=3.274, SD=0.390) revealed no significant differences between groups, t(112)=.804, p=.423. For QI-Evaluation Skills, comparison of scores on the iCoach for teachers (M=3.522, SD=0.513) and coaches (M=3.283, SD=0.378) revealed no significant differences between groups, t(112)=1.362, p=.176. For BM-Goal Formulation Skills, comparison of scores on the iCoach for teachers (M=3.509, SD=0.513) and coaches (M=3.443, SD=0.312) revealed no significant differences between groups, t(12.101)=0.570, p=.579. For BM-Implementation Support Skills, comparison of scores on the iCoach for teachers (M=3.446, SD=0.566) and coaches (M=3.265, SD=0.278) revealed no

significant differences between groups, t(14.562)=1.685, p=.113. For BM-Evaluation Skills, comparison of scores on the iCoach for teachers (M=3.512, SD=0.551) and coaches (M=3.439, SD=0.399) revealed no significant differences between groups, t(112)=0.454, p=.651. For RLC-Goal Formulation Skills, comparison of scores on the iCoach for teachers (M=3.461, SD=0.559) and coaches (M=3.175, SD=0.457) revealed no significant differences between groups, t(107)=1.491, p=.139. For RLC-Implementation Support Skills, comparison of scores on the iCoach for teachers (M=3.465, SD=0.568) and coaches (M=3.011, SD=0.661) revealed significant differences between groups, t(104)=2.147, p=.034. Finally, for RLC-Evaluation Skills, comparison of scores on the iCoach for teachers (M=3.419, SD=0.604) and coaches (M=2.889, SD=1.042) revealed significant differences between groups, t(103)=2.124, p=.036.

At the individual coach level, scale score comparisons varied among teachers and coach (See Table 3). For example, descriptive statistics suggest that coaches A and B rated themselves more effective in general than the teachers they coached. In contrast, coaches C, D, E, H, and I rated themselves less effective in general than their teachers. For Coach F, she rated herself higher for BM, the same for QI, and lower for RLC scale compared to her assigned teachers' ratings of her. Coach G rated herself higher for the scales of BM and RLC and the same for QI compared to her teachers. Please see Table 3 for additional details.

Discussion

The iCoach Assessment System, an on-line, multi-rater, research-based tool, is designed to be a formative assessment for evaluating instructional coaches' competency level and implementation skills. The present investigation examined the initial reliability and validity of the iCoach (pilot version), as well as compared the teacher ratings of the instructional coaches with the instructional coaches own self-ratings. Analyses included examination of internal consistency estimates, item-to-total correlations, freedom from item bias, and the factor structure of the three coaching outcomes (i.e., QI, BM, and RLC) using factor analytic (specifically PCA) approaches. Overall, results suggest that the iCoach has evidence of reliability and validity and that the teachers rated the instructional coaches slightly more favorably than the instructional coaches rated themselves. These findings constitute an important first step toward establishing the iCoach Assessment System as an evidence-based measure for evaluating instructional coaches. Each of the major findings from this investigation is discussed.

Across the three scales, comparisons (i.e., Means and Standard Deviations) between the teacher ratings of the instructional coaches and the instructional coaches own self-ratings demonstrate that overall, the 105 teacher sample rated the instructional coaches slightly more favorably than the instructional coaches rated themselves. With the exception of QI-Goal Formulation Skills, RLC-Implementation Support Skills, and RLC-Evaluation Skills, the difference in ratings was minute (not statistically significant) and thus, the teachers and instructional coaches' ratings were found to be similar. This supports the iCoach's ability to accurately capture the performance of the instructional coach and also establishes the iCoach as an effective multidimensional assessment.

Psychometrics of the iCoach Assessment System

The iCoach is highly internally consistent as evidenced by the Cronbach's alpha for the Total and the three subscales scores. The total and three subscale scores yielded Cronbach's alphas in the range ($\alpha > .90$) that is usually considered acceptable. Item-to-total correlations were computed for the three subscales and were well above Obermiller and Spangenberg's (1998) standards, which recommend that items have item-to-total correlations above .50.

Item bias is essential for measurement development, as it determines whether membership in various groups is likely to systematically influence the relationship between answers and the underlying construct. In the teacher sample of 105, the items from the scales of QI, BM, and RLC were assessed for bias based on the teachers' age, years of experience, and educational degree. The items did not demonstrate item bias based on these three variables, which indicate that items functioned similarly regardless of teacher age, teaching experience, or educational degree. These findings support that coaching outcome scores reflect teachers' ratings of instructional coaches, independent from their own demographic variable influence.

The sample of 105 teachers was used to perform a PCA within the scales of QI, BM, and RLC. Typically, the general rule is that a sample should be at least 10 times the number of variables in order to perform an exploratory factor analysis (EFA; Zaiontz, 2013). Thus, as QI had 53 items, a sample of at least 530 teachers would have been better suited for a PCA. In addition to a small sample size, there was also a linear dependency among variables. Because of this, the results of the PCAs should be interpreted with caution. However, the results can be used as a first step to better refine the iCoach's scales and items. The hypothesized 3-factor model appears to be supported by the results of the PCAs. For QI, BM, and RLC, the 3-factor model accounted for 77.75%, 83.67%, and 87.33% of the cumulative percent of total variance.

Some of the items from the three subscales loaded on multiple factors. This could be due to the teachers rating the instructional coaches similarly across items (e.g., circling "3" on each question). Using a cutoff of .550, the data from the PCAs suggest that many of the items loaded on one factor and that these loadings, based on Comrey and Lee's (1992) interpretation guidelines, were good to excellent. The items from QI seemed to yield the most promising results as 47 of the 53 items loaded on one factor. For BM and RLC, 41 of the 49 items and 29 of the 38 items loaded on one factor. Examining the factor loadings from the scales across the three factors reveals that in QI and BM, component 2 represents Goal Formulation Skills, component 3 represents Implementation Support Skills, and component 1 represents Evaluation Skills. However, for RLC, while component 1 somewhat captures Goal Formulation Skills and Evaluation Skills is clearly represented by component 2, Implementation Support Skills is not differentiated by any of the three components. Taken together, these results may suggest that the amount of items should be decreased across scales for future studies.

Comparison to Other Coaching Assessments

It is useful to contextualize the results of this study by comparing the iCoach to other instruments used to assess instructional coaches in education. Based on the literature (i.e., 2 assessments) and internet search of available instructional coaching assessments (i.e., 24 assessments), only five assessments provided any psychometric evidence. These five coaching assessments were compared to the proposed iCoach assessment instrument in terms of reliability and validity.

Literacy Coach Appraisal Instrument. The literature search resulted in one study from a peer review journal that reported on a Literacy Coach Appraisal Instrument, which included a "Long Form" and "Short Form" version (Lane et al., 2013). The instrument from Lane et al.

(2013) is specifically designed for Literacy Coaches, while the iCoach is content neutral and thus, can be applied to mathematics, science, or literacy coaches. The Literacy Coach Appraisal Instrument had 73 "Literacy Experts" rate items on a 4-point Likert scale as either *Not Important, Minor Importance, Important,* or *Essential*. The resulting 49 items had a majority of the "Literacy Experts" rate it as either *Important* or *Essential.* This is much different than the data collected in this study. As opposed to Literacy Coach Appraisal Instrument, the iCoach was applied to practicing teachers and instructional coaches. The iCoach is more comprehensive as it contains 140 items compared to the Literacy Coach Appraisal Instrument's Long Form, which has 49 items. Also, the Literacy Coach Appraisal Instrument evaluates the literacy coach on how he or she uses technology, serves as a liaison between teacher and administration, and maintains professionalism. The iCoach evaluates the instructional coach mainly on how he or she assists the teacher.

There are similarities between the iCoach and the Literacy Coach Appraisal Instrument. Both instruments are designed to be used as a formative assessment. The Literacy Coach Appraisal Instrument has a section for *Curriculum* (e.g., "How does the literacy coach support the curriculum?"), which is similar to the coaching outcome of QI in the iCoach. In addition, the section for *Teachers* (e.g., "How does the literacy coach support and assist teachers?") is comparable to the coaching outcome of BM and the sections for *Staff Development* (e.g., "How does the literacy coach ensure that staff development needs are met?") and *Resource Management* (e.g., "How does the literacy coach manage literacy resources?") can be thought of as somewhat similar to the coaching outcome of RLC. Like the iCoach, the literacy coach's ability to work with the teacher to collect accurate assessment data is measured in the Literacy Coach Appraisal Instrument.

Coaching Evaluation Survey – Revised. One of the instruments that resulted from an internet search using *Google* was the Coaching Evaluation Survey – Revised from the Florida's Problem Solving/Response to Intervention (PS/RtI) Project (Castillo et al., 2013). Similar to the iCoach Assessment System, this instrument is content neutral. It has 20 items which are declarative statements (e.g., "My school's PS/RtI coach is an effective listener) and are rated on a 5-point Likert scale as either *Strongly Disagree, Disagree, Neutral, Agree,* or *Strongly Agree.* There is also an option for *Do Not Know.* It is intended for the School-Based Leadership Team (SBLT) to complete, which is comprised of approximately six to eight staff members who spearhead implementation of PS/RtI in a school (Castillo et al., 2013). However, Castillo et al. (2013) note that the instrument can be used for the coaches to rate themselves. This instrument is both a formative and summative assessment tool. It is not designed for performance evaluation, but solely to inform the coaches' professional development. The Coaching Evaluation Survey – Revised consists of three factors: *Role, Function, and Activities of the PS/RtI Coach; Modeling of the Problem Solving Process;* and *Interpersonal/Communication Skills*. In examining the items of this instrument, it is apparent that numerous items resemble the iCoach coaching skills of Implementation Support Skills and Evaluation Skills and the coaching action of Identifying Needs and Resources.

The Coaching Evaluation Survey – Revised has much supporting psychometric evidence; however, it has not yet been published in a peer-reviewed journal. This instrument has evidence of content validity, construct validity, and internal consistency. It has construct validity as evidenced by both an EFA and Confirmatory Factor Analytic (CFA) analyses. The EFA, which used the Promax method (i.e., oblique rotation), was conducted on a sample of 506 SBLT members and the three factors accounted for 96% of the common variance. The factor loadings

were all greater than .45 and none of the items loaded on multiple factors. This is in contrast to the iCoach Assessment System, where some of the items loaded on multiple factors. Similar to the iCoach, the Coach Evaluation Survey – Revised had high levels of internal consistency as measured by Cronbach's alpha (i.e., *Role, Function, and Activities of the PS/RtI Coach* α=.97, *Modeling of the Problem Solving Process* α=.97, and *Interpersonal/Communication Skills* α=.96).

EMC Teacher Reflection and Impact Survey (TRIS). The TRIS, which was found through a *Google* search, is used for teachers to evaluate their mathematics coaches (Yopp, Burroughs, & Sutton, 2010b). It consists of 34 items with over 3 scales (i.e., *Topics Discussed*, *Coaching Relationships,* and *Impact of Coaching*). The *Topics Discussed* and *Coaching Relationships* scales are considered reflection items and the *Impact of Coaching* scale is categorized as consisting of impact items. The *Topics Discussed* scale presents statements regarding which topics the teacher discussed with the mathematics coach and then the teacher provides ratings on a 5-point Likert scale where 1 is *Not at All* and 5 is *To a Great Extent.* This scale consists of sections concerning mathematics content, mathematical concept and inquiry, classroom environment/culture, and reflection and planning. The *Coaching Relationships* scale presents statements regarding the teacher's communication, comfort level, and feeling valued and respected by the mathematics coach and then the teacher selects ratings on the 5-point Likert scale. Similar to the *Topics Discussed* scale, the *Impact of Coaching* scale also presents statements regarding which topics the teacher discussed with the mathematics coach (e.g., "Discussions with my coach about ways to infuse more conceptual understandings into my lessons"), but then the teacher has to provide ratings based on the impact it had on them (i.e., 6-point Likert scale where 0 is *Didn't Discuss or not a Topic of Emphasis*, 1 is *Discussed, but no*

impact, 3 is *Moderate Impact*, and 5 is *Very Large Impact*). Some of the items relate to the iCoach coaching outcomes of QI and BM, the coaching skill of Evaluation Skills, and the coaching action of Setting Goals. This instrument is not content neutral and appears to be a summative evaluation tool.

An EFA with maximum likelihood extractions and varimax rotations were conducted on the reflection items, as well as the impact items for a sample of 173 teachers. This data was not published in a peer-reviewed publication. For the reflection items, the resulting data suggests a 2-factor solution (i.e., *Topics Discussed* and *Coaching Relationships* scales) that explains 73.33% of the total variance. The factor loadings are all greater than .67 and only two of the items loaded on both factors (the secondary loadings were .slightly above .4). For the impact items, the resulting data suggests a single factor solution (i.e., *Impact of Coaching* scale) that explains 69.42% of the total variance. The factor loadings were all greater than .80. Compared to the iCoach, the items on the TRIS seem to differentiate factors better, but the 3-factor solution on the iCoach accounts for more of the total variance (73.33% and 69.42% compared to 77.75%, 83.67%, and 87.33%). The iCoach and TRIS both have high levels of internal consistency as measured by Cronbach's alpha (i.e, *Topics Discussed* α=.973, *Coaching Relationships* α=.953, and *Impact of Coaching* α=.967).

EMC Coaching Skills Inventory (CSI). The CSI, which was also found though a *Google* search, measures a mathematics coach's self-efficacy with various coaching abilities (Yopp, Burroughs, & Sutton, 2010a). The instrument has 20 items across five categories of coach/teacher relationships, coaching skills, mathematics content, mathematics-specific pedagogy, and general pedagogy. The items are measured on a 5-point Likert scale where 1 is *Not at All Effective* and 5 is *Very Effective*. Most of the items on this measure can be considered

similar to the iCoach coaching outcome of QI. This instrument is not content neutral and appears to have utility as both a formative and summative evaluation tool.

An EFA with maximum likelihood extractions and varimax rotations was conducted on the items for a sample of 57 mathematics coaches. This evidence was not published in a peer-reviewed publication. Since the sample was modest, the results of this EFA should be interpreted with caution. The resulting data suggests a 3-factor solution of *Mathematics Content and Mathematics Specific Pedagogy, Student Centered Pedagogy Coaching,* and *Building Coaching Relationships.* The 3 factors explain 62.80% of the total variance. This is in contrast to the iCoach, where the 3-factor solution accounted for more of the total variance within scale (62.80% compared to 77.75%, 83.67%, and 87.33%). The factor loadings were all greater than .56 and it does not appear that any of the items loaded on multiple factors. Like the iCoach, the CSI has high levels of internal consistency as measured by Cronbach's alpha (i.e., *Mathematics Content and Mathematics Specific Pedagogy* α=.935, *Student Centered Pedagogy Coaching* α=.932, and *Building Coaching Relationships* α=.822).

Implications

Teacher evaluation practices have been scrutinized and reform has been prioritized in schools across the nation. With the authorization of the ESSA, state education departments now have more flexibility in devising teacher evaluation models (Sawchuk, 2016). As teacher evaluation practices change, professional development becomes essential for teacher success. Instructional coaching is a widespread professional development strategy that is currently being implemented in schools. To be an effective instructional coach, it is important that ongoing performance assessment and data-based feedback is given. An empirically validated instructional coaching assessment that can identify instructional coaches' areas of strength and areas needing

improvement is urgently needed to ensure that teachers are being provided with necessary levels of support and guidance.

The iCoach Assessment System is designed to evaluate instructional coaches' competency level and implementation skills. It contains three coaching outcomes, three coaching skills, and six coaching actions. In this study, we presented initial reliability and validity evidence for the iCoach. The iCoach appears to be highly internally consistent, have freedom from item bias based on group membership, and each of the three coaching outcomes appears to have a 3-factor structure. As few reliable and valid assessments exist for instructional coaches, the iCoach can be used in schools in order to provide continuous performance feedback that assists school leaders in developing the skills of instructional coaches. The iCoach's use as a formative assessment tool is supported by its strong internal consistency across and within scales as well as its design as an evidence-centered assessment. Since the iCoach adheres to an evidence-centered assessment design, the coaches are observed and their actions are rated. This should theoretically lead to the iCoach being sensitive to and readily capturing improved coaching practices throughout the school year. In addition, the iCoach gives school leaders a tool that is comprehensive, both conceptually and in length (currently has 140 items), and content neutral and thus, can be used with instructional coaches from any subject (e.g., literacy, mathematics, etc.). The iCoach is a 360 evaluation, which ensures that it will incorporate each stakeholder's perspective and produce an accurate view of the instructional coach's performance. If instructional coaches can receive accurate assessment data and feedback, then their skills and competencies can be improved upon, which in turn, ensures that they are effective in meeting teacher needs.

Limitations and Future Research

The current study includes limitations. First, the study had a modest sample size than was ideal for a PCA. As the sample came from five charter schools, the instructional coach sample was entirely female, and the teacher sample taught Pre-K through 8th grade and not high school, it may be hard to generalize the findings of this study to other school communities in the nation. Furthermore, certain items on each of the iCoach Scales were not able to be rated by teachers or coaches. For these items, both groups indicated that there was no opportunity to engage in those constructs. Although these items were excluded from analyses, this reduced the overall amount of items available for rating individual coach quality. In addition, no independent observations of coaches' behavior were conducted for this study and coaching session audio tapes were not collected and subsequently reviewed by independent coders. Therefore, the quality of coaching as measured by direct observation and post session analysis is unknown at this time. Another limitation is that the dosage of coaching provided to teachers likely varied across instructional coaches and this might have influenced the ratings given. Finally, many of the instructional coaches previously worked as teachers in the schools where they were now coaching and thus, had already established relationships with teachers. This might have influenced the ratings.

In future research of the iCoach, it would be useful to compare a complete 360 evaluation which would involve not only the teacher and instructional coach ratings, but also the administrators' scoring of instructional coaches. Using a complete 360 evaluation would allow researchers to determine if adding another perspective (i.e., administrators) affected the ratings of the instructional coaches (e.g., are the administrators' rating different than the teachers?) and whether this impacted the data that was collected. Future research should also examine the utility of the iCoach in repeated measurements for individual coaches in order to determine whether the

iCoach is sensitive to change within as well as between coaches. In addition, future research should examine the factor structure of the iCoach with a larger sample. A larger sample should lead to more variability in scores, which would likely diminish the possibility of linear dependencies among items. This would make the data more conducive to factor analytic approaches. Future research that contained a more heterogeneous sample of teachers and coaches who work in a variety of settings (e.g., public schools) would lead to increased generalizability of results. As mentioned above, studies that included independent observations of coaches' behavior and measured the dosage of coaching would provide further insight into the quality of the coaching as well as eliminate potential confounds.

Conclusion

As instructional coaching is an increasingly prevalent form of PD being used in schools across the nation, it is important that a formative assessment tool exists for evaluating and developing instructional coaches. Such a tool would ensure that the instructional coaches are effective in meetings the needs of the students and teachers. The iCoach Assessment System offers a practical and promising tool for assessing instructional coaches. The initial reliability and validity evidence presented in this study indicates that the iCoach has promise as a psychometrically sound measure for assessing instructional coaches' competencies and skills.

References

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.

Andrews, D., & Lewis, M. (2007). Transforming practice from within: The power of the professional learning community. In L. Stoll & K.S. Louis (Eds.), *Professional learning communities: Divergence, depth, and dilemmas.* Maidenhead, U.K.: Open University Press.

*Arkansas Department of Education. (2014). Instructional Specialist Framework. In *Arkansas TESS and LEADS Resources* (pages 233-246). Retrieved from http://www.arkansased.gov/public/userfiles/HR_and_Educator_Effectiveness/TESS/Handbook6-80968.0.pdf

Beglau, M., Hare, J.C., Foltos, L., Gann, K., James, J., Jobe, H., … Smith, B. (2011). *Technology, Coaching, and Community.* Retrieved from http://www.isteconference.org/uploads/ISTE2013/HANDOUTS/KEY_81724011/Coaching_Whitepaper_digital.pdf

Berliner, D.C. (1986). In Pursuit of the Expert Pedagogue. *Educational Researcher, 15*(7), 5-13.

Bohrnstedt, G.W. (1969). A Quick Method for Determining the Reliability and Validity of Multiple-Item Scales. *American Sociological Review, 34*(4), 542-548.

Briere, D. E., Simonsen, B., Sugai, G., & Myers, D. (2015). Increasing new teachers' specific praise using a within-school consultation intervention. Journal of Positive Behavior Interventions, 17(1), 50-60.

Browers, A. & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*(2), 239-253.

Campbell, P.F., & Malkus, N.N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111*(3), 430-454.

*Castillo, J.M., Batsche, G.M., Curtis, M.J., Stockslager, K., March, A., Minch, D., & Hines, C. (2013). Coaching Evaluation Survey – Revised. In *Problem Solving/Response to Intervention evaluation tool technical assistance manual - revised* (Chapter Three Tools for Examining Infrastructure Development). Retrieved from http://floridarti.usf.edu/resources/program_evaluation/evaluation_tools/index.html

*Center Grove Community School Corporation. (n.d). *Instructional Specialist Effectiveness Rubric*. Retrieved from http://www.centergrove.k12.in.us/staffevaluations

Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act.* Retrieved from http://www.cep-dc.org/publications/index.cfm?selectedYear=2006

Cicchetti, D.V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment, 6*(4), 284-290.

*Clark Pleasant Community Schools. (n.d). *Instructional Coach Evaluation Rubric*. Retrieved from http://www.cpcsc.k12.in.us/cpcsc/media/evalres/instructional_coach_rubric.pdf

Commission on Effective Teachers and Teaching. (2012). *Transforming Teaching: Connecting Professional Responsibility With Student Learning- A Report to the National Education Agency*. Retrieved from http://www.nea.org/assets/docs/Transformingteaching2012.pdf

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

Cook, J.D. (2014). *The Global Relational Attitudes Conflict Exam (GRACE) Scale: An*

Exploratory Factor Analysis (Doctoral dissertation). Retrieved from ProQuest. (UMI 3579797).

Cornett, J., & Knight, J. (2009). Research on coaching. In J. Knight (Ed.), *Coaching: Approaches and perspectives* (pp. 192-216). Thousand Oaks, CA: Corwin Press, Inc.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life.* New York, NY: Basic Books.

*Danielson, C. (2012). *Framework for Instructional Specialist Evaluation Instrument 2013.* Retrieved from https://danielsongroup.org/framework/

Denton, C.A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation, 19*(2), 150-175.

Dillon, S. (2008). Head of teachers' union offers to talk on tenure and merit pay. *The New York Times.* Retrieved from http://www.nytimes.com/2008/11/18/education/18teach.html.

Duncan, A. (2009). Interview with Wolf Blitzer, Transcript. The Situation Room. Retrieved from http://transcripts.cnn.com/TRANSCRIPTS/0903/10/sitroom.02.html.

Duncan, A. (2012). The Tennessee Story. *The Huffington Post.* Retrieved from http://www.huffingtonpost.com/arne-duncan/the-tennessee-story_b_1695467.html

Emmer, E.T., & Stough, L.M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational psychologist, 36*(2), 103-112.

Espin, C.A., & Yell, M.L. (1994). Critical indicators of effective teaching for preservice teachers: Relationships between teaching behaviors and ratings of effectiveness. *Teacher Education and Special Education, 17,* 154-169.

*Friendship Public Charter School. (2010). *Coach Evaluation Performance Management Rubric*. Retrieved from http://tntp.org/assets/tools/FPCS_Coaching_Model_at_Friendship_Public_Charter_School_TSLT_0311.pdf

Goldstein, J. (2004). Making sense of distributed leadership: The case of Peer Assistance and Review. *Educational Evaluation and Policy Analysis* 26 (2): 173-97.

Goldstein, J. (2007). Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. *American Journal of Education, 113,* 479-508.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). Multivariate data analysis 7th Edition. Upper Saddle River, NJ: Pearson Education.

Hershfeldt, P. A., Pell, K., Sechrest, R., Pas, E. T., & Bradshaw, C. P. (2012). Lessons learned coaching teachers in behavior management: The PBIS plus coaching model. *Journal of Educational and Psychological Consultation*, *22*(4), 280-299.

International Reading Association, National Council of Teachers of English, National Council of Teachers of Mathematics, National Science Teachers Association, & National Council for the Social Studies (2006). *Standards for middle and high school literacy coaches.* Newark, DE: Author.

Reddy, L.A., Glover, T., Elliott, S., & Kurz, A. (2016). iCoach Assessment System. Westfield, NJ: iCoach Learning Innovations.

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

Knight, J. (2006). Eight factors for realizing better classroom teaching through support, feedback, and intensive, individualized professional learning. *The School Administrator,*

63(4), 18 - 25.

*Knight, J. (2013). Tool 8: Knowledge of Practices, Tool 9: Partnership Communication Self-Assessment, and Tool 10: Partnership Leadership Self-Assessment. In *Toolkit Coaching Coaches* (pages 11-13). Retrieved from http://www.instructionalcoaching.com/resources/pdfs/ICG_CoachCoaches_Tools.pdf

Koppich, J. (2005). Addressing teacher quality through induction, professional compensation, and evaluation: The effects on labor-management relations. *Educational Policy, 19*(1): 90-111.

Kowal, J., & Steiner, L. (2007). *Instructional Coaching*. Retrieved from http://files.eric.ed.gov/fulltext/ED499253.pdf

*Lane, M., Robbins, M., & Price, D. (2013). Establishing Content Validity for a Literacy Coach Performance Appraisal Instrument. *Literacy Research and Instruction, 52*(3), 230-253.

*Langrange Area Department of Special Education. (2014). *Instructional Specialist Evaluation Rubric.* Retrieved from http://www.ladse.org/wp-content/uploads/2014/08/Instructional-Specialist-14-15-Revised-07-28-14.pdf

Lee, S.C., DeChenne, S.E., Nugent, G., Kunz, G.M., & Houston, J. (2014). *Case Study: Impact of Instructional Coaching on Science Teachers' Inquiry-Based Instruction in Rural Schools.* (R^2Ed Working Paper No. 2014-13). Retrieved from the National Center for Research on Rural Education: r2ed.unl.edu.

*Literacy Coaching Clearinghouse. (2009). *Self-Assessment for Elementary Literacy Coaches.* Retrieved from http://www.literacycoachingonline.org/briefs/tools/self_assessment_for_elem_lit_coaches.pdf

*Marion County. (2013). *Marion County Evaluation Student Support Services Personnel Rubrics.* Retrieved from http://manateeea.fea.aft.org/files/mcies_student_support_services_personnel_rubrics_8_2012.pdf

Marsh, J.A., McCombs, J.S., & Martorell, F. (2010). How Instructional Coaches Support Data-Driven Decision Making Policy Implementation and Effects in Florida Middle Schools. *Educational Policy, 24*(6), 872-907. doi: 10.1177/0895904809341467

*Marzano, R., Carbaugh, B., & Toth, M. (2013). *Marzano Center Non-Classroom Instructional Support Member Evaluation Form*. Retrieved from http://www.marzanocenter.com

Mead, S. (2012). *Recent state action on teacher Effectiveness: What's in state laws and regulations?* Retrieved from http://bellwetheredecuation.org/sites/default/files/RSA-Teacher-Effectiveness.pdf

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

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. CRESST Technical Paper Series. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Mislevy, R. J., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9)*. Menlo Park, CA: SRI International.

National Education Association. (2011). *Promoting and Implementing the National Education Association Policy Statement on Teacher Evaluation and Accountability: An NEA Toolkit.* Retrieved from http://www.nea.org/assets/docs/2011NEA_Teacher_Eval_Toolkit.pdf.

Nelson, T.H. (2009). Teachers' collaborative inquiry and professional growth: Should we be optimistic? *Science Education, 93*(3), 548-580.

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.

*North Bergen School District. (2014). *Formal Observation/Evaluation Rubric – Literacy Coach.* Retrieved from http://www.northbergen.k12.nj.us/site/default.aspx?PageID=1

Obama, B. (2009). White House "Open for Questions" Town Hall Transcript. *The New York Times*. Retrieved from http://www.nytimes.com/2009/03/26/us/politics/26text-obama.html?pagewanted=all.

Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of consumer psychology*, *7*(2), 159-186.

*Okaloosa School District. (n.d.). *Formal Observation/Evaluation Rubric – Literacy Coaches*. Retrieved from http://www.okaloosaschools.com/content/hr-forms

Oliver, R. M., & Reschly, D. J. (2007). *Effective classroom management: Teacher preparation and professional development*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/topics/effectiveClassroomManagement.pdf

*Orange County School. (2009). *Literacy Coach Observation Data Instrument.* Retrieved from https://www.orange.k12.nc.us/HR/HR_links/HR_Docs/HR_Lit_Coach_End_YR.pdf

Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, *26*(5), 696-729.

*Parkway School District. (2011). *Professional Learning and Evaluation Plan – Secondary Literacy Coach*. Retrieved from http://parkwayschools.net/inside/curriculum/CA/secondary/file/MS%20Literacy%20Coach%20Evaluation.pdf

*Pennsylvania Department of Education. (n.d.) *Possible examples of how the Framework for Teaching could apply to Instructional Coaches.* Retrieved from http://www.pdesas.org/Instruction/Frameworks

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*(4), 431-451.

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*(2), 299-312.

Reddy, L.A., Dudek, C.M., Kettler, R., Kurz, & Peters, S. (2016). Measuring Educators' Attitudes and Beliefs About Evaluation: Construct Validity and Reliability of the Teacher Evaluation Experience Scale. *Educational Assessment, 21*(2), 120-134.

Reddy, L.A., Glover, T.A., Kurz, A., & Elliot, S.N. (2015). *iCoach Assessment System.* iCoach Learning Innovations LLC, Westfield, NJ.

Resnick, R. (2010). *National survey of professional development trends: 2010*. Education Market Research, Rockaway Park, NY.

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: Wiley.

Reynolds, C. R., Willson, V. L., & Chatman, S.R. (1984). Item bias on the 1981 revisions of the Peabody Picture Vocabulary Test using a new method of detecting bias. *Journal of Psychoeducational Assessment, 2*(3), 219–224. doi:10.1177/073428298400200306

*Rockwood School District. (2009). *Instructional Coach Evaluation Instruments.* Retrieved from http://www.rockwood.k12.mo.us/boardofeducation/policiesandregulations/District%20Forms/Instructional%20Coach%20-%20Evaluation%20Handbook.pdf

Rush, L.S., & Young, S. (2011). Wyoming's Instructional Facilitator Program: Teachers' Beliefs about the Impact of Coaching on Practice. *Rural Educator, 32*(2), 13-22.

Sanetti, L. M. H., 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*(4), 454-465.

Sawchuk, S. (2016). ESSA Loosens Reins on Teacher Evaluations, Qualifications. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html

*Shelby County School. (2013). *Teacher Effectiveness Measure Framework Instructional Coaches.* Retrieved from http://www.scsk12.org/uf/TalentManagement/PDFs/TEM3_0_Coaches_Rubric-FINAL7_1.pdf

Shernoff, E.S., Lakind, D., Frazier, S.L., & Jakobsons, L. (2015). Coaching early career teachers in urban elementary school: A mixed-method study. *School Mental Health, 7*(1), 6-20.

Shute, V., Kim, Y.J., & Razzouk, R. (2013). *ECD for Dummies*. Retrieved from http://myweb.fsu.edu/vshute/ECD%20for%20Dummies/ECD%20for%20Dummies.swf

*Springfield Public Schools. (2014). *Framework of Professional Practice for Instructional*

Coaches. Retrieved from http://www.springfieldpublicschoolsmo.org/

*St. Charles School District. (2014). *St. Charles School District Instructional Coach Growth Guide and Evaluation Documents*. Retrieved from http://www.stcharles.k12.mo.us/downloads/HR/Instructional_Coach_Evaluation_MARCH_2014-A.pdf

Sternberg, R.J., Torff, B., & Grigorenko, E.L. (1998). Teaching triarchically improves school achievement. *Journal of Educational Psychology, 90*(3), 374-384.

*St. Joseph School District. (n.d.) *Criteria for Effective School-Based Coaching.* Retrieved from http://www.sjsd.k12.mo.us/cms/lib3/MO01001773/Centricity/Domain/57/Coach%20Eval%20form.doc

Stormont, M., Reinke, W. M., Newcomer, L., Marchese, D., & Lewis, C. (2015). Coaching Teachers' Use of Social Behavior Interventions to Improve Children's Outcomes A Review of the Literature. *Journal of Positive Behavior Interventions*, *17*(2), 69-82.

Sutherland, K.S., & Wehby, J.H. (2001). Exploring the Relationship Between Increased Opportunities to Respond to Academic Requests and the Academic and Behavioral Outcomes of Students with EBD A Review. *Remedial and Special Education, 22*(2), 113-121.

Toch, T. & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education.* Washington, DC: Education Sector.

U.S. Department of Education, Institute of Education Sciences. (2014). *A summary of Professional Development Research: National Center for Special Education Research FY 2006 – FY 2014*. Retrieved from http://ies.ed.gov/ncser/pdf/PD_2014.pdf

U.S. Department of Education, Institute of Education Sciences. (2016). *Request for Applications*

Education Research Grants CFDA Number: 84.305A. Retrieved from http://ies.ed.gov/funding/pdf/2017_84305A.pdf

Utley, C. A. (2011). A Psychometric Investigation of the Multicultural and Special Education Survey: An Exploratory Factor Analysis. *Learning Disabilities: A Contemporary Journal*, *9*(1), 47-70.

Vernon-Feagans, L., Kainz, K.L., Hedrick, A., Ginsberg, M.C., & Amendum, S.J. (2013). Live webcam coaching to help elementary classroom teachers provide effective literacy instruction for struggling readers. The Targeted Reading Intervention. *Journal of Educational Psychology*, 105(4), 1175-1187.

Vygotsky, L.S. (1986). *Thought and language, revised edition.* (A. Kozulin, Ed.). Cambridge, MA: MIT Press.

*Washakie County School District. (2011). *School-Based Instructional Facilitators – District Instructional Facilitator Plan*. Retrieved from http://www.wsh1.k12.wy.us

Winerip, M. (2011). Helping Teachers Help Themselves. *The New York Times.* Retrieved from http://www.nytimes.com/2011/06/06/education/06oneducation.html

Wren, S., & Vallejo, D. (2009). *Effective collaboration between instructional coaches and principals.* Retrieved from http://www.balancedreading.com/Wren_&_Vallejo_Coach_Principal_Relatinships.pdf

*Yopp, D., Burroughs, E., & Sutton, J. (2010a). *Coaching Skills Inventory*. Bozeman, MT, and Denver, CO: Examining Mathematics Coaching (Montana State University and RMC Research Corporation). Supported by NSF Discovery Research K-12 Program, Award No. 0918326.

*Yopp, D., Burroughs, E., & Sutton, J. (2010b). *Teacher Reflection and Impact Survey*.

Bozeman, MT, and Denver, CO: Examining Mathematics Coaching (Montana State University and RMC Research Corporation). Supported by NSF Discovery Research K-12 Program, Award No. 0918326.

Zaiontz, C. (2013). *Validity of Correlation Matrix and Sample Size.* Retrieved from http://www.real-statistics.com/multivariate-statistics/factor-analysis/validity-of-correlation-matrix-and-sample-size/

Zillmer, E. & Vuz, J. (1995). Factor analysis with Rorschach data. In J. Exner, Jr. (Ed.), *Methods and issues in Rorschach research* (pp. 251-306). Hillsdale, NJ: Lawrence Erlbaum Associates.

Table 1

Assessment Instruments for Instructional Coaches (N=26)

Source	Year	Title	Informant	Scale Description	Domains
Center Grove Community School Corp.	N/A	Instructional Specialist	School Administration	Effective or Highly Effective	Planning & Preparation; The Environment; Delivery of Service; Professional Responsibilities
Clark Pleasant Community School in IN	N/A	Instructional Coach	School Administration	Ineffective, Improvement Necessary, Effective, or Highly Effective	Professional Relationship; Effective Coaching Skills; The Coaching Program
Okaloosa School District in FL	N/A	Literacy Coach	School Administration	Unsatisfactory, Improvement Needed/Developing, Effective, or Highly Effective	Professional Knowledge & Planning; Context of Learning; Professional Development & Coaching Activities; Professional Responsibilities.
PA Dept. of Education	N/A	Instructional Coach	School Administration	Failing, Needs Improvement, Proficient, or Distinguished	Planning & Preparation; Environment; Service Delivery; Professional Development/Professional Responsibilities
St. Joseph School District in MO	N/A	School-Based Coach	Self-assessment	Novice, Developing, or Accomplished	Demonstrated skill as a classroom teacher; Relationship Building; Skilled Facilitation; Data-Driven Coaching; Adult Learning; Learning Stance; Time Management; Reflective Dialogue; Productive Relationship with the School Leadership.
Literacy Coaching Clearinghouse	2009	Literacy Coach	Self-assessment	6-point scale (Please see article for more details)	Foundations of Literacy; Assessment; Content Area; Instruction; Writing; Differentiated Instruction; Classroom Coaching; Facilitating Adult Learning; Building Capacity Within the School; & Working Within a Broader School Reform Context.
Orange County School in NC	2009	Literacy Coach	School Administration	Not Demonstrated, Developing, Proficient, Accomplished, or Distinguished	Planning & Facilitating Teaching & Learning; Planning & Facilitating Information Access & Delivery, Evaluation, & Use; Planning & Facilitating Program Administration.

Table 1 Continued

Rockwood School District in MO	2009	Instructional Coach	School Administration	Does Not Meet the Standard, Partially/Approaching the Standard, Meets the Standard, or Exceeds the Scale	Planning & Preparation; Instruction; Environment; Professionalism
Friendship Public Charter School in Washington, D.C.	2010	Instructional Performance Coach	Self-assessment	Unsatisfactory, Needs Improvement, Proficient, or Exemplary	Excellent Teaching & Learning; Outstanding Leadership; Environment Conducive to Learning
Yopp, Burroughs, & Sutton, 2010 from Examining Mathematics Coaching (EMC Coaching Skills Inventory)[a]	2010	Mathematics Instructional Coach	Self-assessment	5-Point Scale on continuum of Not At All Effective to Very Effective	Coach/Teacher Relationships; Coaching Skills; Mathematics Content; Mathematics-Specific Pedagogy; General Pedagogy
Yopp, Burroughs, & Sutton, 2010 from Examining Mathematics Coaching (EMC Teacher Reflection & Impact Survey)[a]	2010	Mathematics Instructional Coach	Teacher	5-Point Scale on continuum of Not At All to Great Extent	Interactions with the Coach; Frequency of Various Activities; Coaching Relationship; Topics Discussed: Mathematics Content, Mathematical Concept & Inquiry, Classroom Environment/Culture; Reflection & Planning; Impact on Teacher Practice
Parkway School District in MO	2011	Literacy Coach	School Administration	Ineffective, Development Area, Effective, or Exemplary	Coaching Process; Curriculum & Instruction; Culture of Learning; Professional Responsibility.
Washakie County School District in WY	2011	Instructional Facilitator	School Administration	Beginning, Emerging, or Established	Role/Responsibility; Professional Competencies
Charlotte Danielson	2012	Instructional Specialist	School Administration	Unsatisfactory, Basic, Proficient, or Distinguished	Planning & Preparation; The Classroom Environment; Instruction; Professional Responsibilities
Castillo, Batsche, Curtis, Stockslager, March, Minch, & Hines, 2013 from Florida's Problem Solving/Response to Intervention (PS/RtI) Project[a]	2013	Problem Solving/ Response to Intervention Coach	School Administration, Teacher, Self-Assessment	Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree, or Do Not Know	Role, Function, and Activities; Modeling of the Problem Solving Process; Interpersonal/Communication Skills

Table 1 Continued

Lane, Robbins, & Price, 2013 (Short Form)[b]	2013	Literacy Coach	School Administration	4-point scale (i.e., "U, BE, P, EE"). Also a choice for N/A.	Resource Management; Staff Development; Assessment; Professional Disposition; Teachers
Lane, Robbins, & Price, 2013 (Long Form)[b]	2013	Literacy Coach	School Administration	4-point scale (i.e., "U, BE, P, EE"). Also a choice for N/A.	Curriculum; Teachers; Staff Development; Technology; Liaison; Assessment; Home & Family Education; Resource Management; Professional Disposition
Marion County Public School in FL	2013	Instructional Support Services	School Administration	Unsatisfactory, Needs Improvement/Developing, Effective, or Highly Effective	Planning & Preparation; The Environment; Delivery of Service; Professional Responsibilities
Robert Marzano	2013	Non-Classroom Instructional Support	School Administration	Not Using, Beginning, Developing, Applying, or Innovating	Instructional Support Strategies & Behaviors; Planning & Preparing for Implementation of Goals & Scaffolding of Content or Activities; Reflecting on Teaching; Collegiality & Professionalism
Shelby County School in TN	2013	Instructional Coach	School Administration	Significantly Below Expectation, Below Expectations, Meeting Expectations, Above Expectations, or Significantly Above Expectations	Supporting Academic Achievement through Teacher Effectiveness; Supporting Academic Achievement through Data Management; Supporting Academic Achievement through Collaboration & Professional Development; Supporting Academic Achievement through District & School-specific Instructional Initiatives; Supporting Academic Achievement by Increasing Content & Coaching Knowledge; Scope of Work; Delivery of Services
Arkansas' Dept. of Education	2014	Instructional Specialist	School Administration	Unsatisfactory, Basic, Proficient, or Distinguished	Planning & Preparation; The Environment; Delivery of Service; Professional Responsibilities
Jim Knight	2014	Instructional Coach	Self-assessment	7-Point Scale on continuum of Disagree to Agree	Knowledge of Practices; Partnership Communication; Partnership Leadership
Lagrange Area Dept. of Special Education in IL	2014	Instructional Specialist	School Administration	Unsatisfactory, Basic, Proficient, or Distinguished	Planning & Preparation; The Environment; Delivery of Service; Professional Responsibilities

Table 1 Continued

North Bergen School District in NJ	2014	Literacy Coach	School Administration	Ineffective, Developing, Effective, or Highly Effective	Professional Knowledge & Planning; Context of Learning; Professional Development & Coaching Activities; Professional Responsibilities.
Springfield Public School in IL	2014	Instructional Coach	School Administration	Unsatisfactory, Basic/Needs Improvement, Proficient, or Distinguished/Excellent	Planning & Preparation; The Environment; Delivery of Service; Professional Responsibilities
St. Charles School District in MO	2014	Instructional Coach	School Administration	Beginning 1, Beginning 2, Beginning 3, Beginning 4, Proficient 5, Proficient 6, or Distinguished 7	Builds & Maintains Positive Relationships; Participates in & Provides Professional Development; Establishes a Culture of Learning, Knowledge of Curriculum & Resources, Planning & Preparation, Data Collection & Analysis; Student Assessment & Data Analysis

a Psychometric data in technical report/manual only

b Psychometric data published in a peer review journal article

Table 2

Sample Characteristics

Characteristic	Teachers (*n* = 105)	Instructional Coaches (*n* = 9)
Age (years)		
Mean (SD)	32.21 (8.69)	36.11 (8.74)
Range	23-59	27-51
Gender		
Male	16 (15.20%)	0 (0.00%)
Female	89 (84.80%)	9 (100.00%)
Ethnicity		
Not Hispanic or Latino	99 (94.30%)	9 (100.00%)
Hispanic or Latino	6 (5.70%)	0 (0.00%)
Race		
Caucasian	81 (77.14%)	5 (55.55%)
Black or African-American	16 (15.24%)	2 (22.22%)
Asian	4 (3.81%)	1 (11.11%)
Other	4 (3.81%)	1 (11.11%)
Degree Level		
Bachelor's	73 (69.50%)	5 (55.55%)
Master's	32 (30.50%)	4 (44.44%)
Years of Teaching Experience		
Mean (SD)	5.4 (5.84)	9.4 (6.02)
Range	0-39	3.5-20
Grade Level Taught/Coached		
Pre-K	1 (1.00%)	0 (0.00%)
Kindergarten	14 (13.30%)	0 (0.00%)
1st	10 (9.50%)	0 (0.00%)
2nd	14 (13.30%)	0 (0.00%)
3rd	7 (6.70%)	0 (0.00%)
4th	7 (6.70%)	0 (0.00%)
5th	5 (5.70%)	0 (0.00%)
6th	3 (2.90%)	0 (0.00%)
7th	4 (3.80%)	0 (0.00%)
8th	2 (1.90%)	0 (0.00%)
9th	0 (0.00%)	0 (0.00%)
10th	0 (0.00%)	0 (0.00%)
11th	0 (0.00%)	0 (0.00%)
12th	0 (0.00%)	0 (0.00%)
Multiple Grades	37 (35.20%)	9 (100.00%)

Table 3

Comparison Scores of Instructional Coaches and Teachers

All Instructional Coaches				**All Teachers**			
	N	**M**	**SD**		**N**	**M**	**SD**
QI	9	3.26	0.32	**QI**	105	3.50	0.50
BM	9	3.39	0.30	**BM**	105	3.50	0.52
RLC	9	3.08	0.56	**RLC**	105	3.45	0.55
Instructional Coach A				**Teachers for Instructional Coach A**			
QI	1	3.81	N/A	**QI**	16	3.21	0.64
BM	1	3.78	N/A	**BM**	16	3.19	0.66
RLC	1	3.92	N/A	**RLC**	16	3.19	0.64
Instructional Coach B				**Teachers for Instructional Coach B**			
QI	1	3.08	N/A	**QI**	6	2.98	0.27
BM	1	3.11	N/A	**BM**	6	3.04	0.26
RLC	1	3.00	N/A	**RLC**	6	3.03	0.26
Instructional Coach C				**Teachers for Instructional Coach C**			
QI	1	2.92	N/A	**QI**	15	3.49	0.43
BM	1	2.96	N/A	**BM**	15	3.53	0.46
RLC	1	2.84	N/A	**RLC**	14	3.47	0.40
Instructional Coach D				**Teachers for Instructional Coach D**			
QI	1	2.96	N/A	**QI**	11	3.29	0.64
BM	1	3.22	N/A	**BM**	11	3.28	0.63
RLC	1	2.92	N/A	**RLC**	11	3.29	0.79
Instructional Coach E				**Teachers for Instructional Coach E**			
QI	1	3.38	N/A	**QI**	11	3.69	0.38
BM	1	3.64	N/A	**BM**	11	3.73	0.33
RLC	1	3.50	N/A	**RLC**	11	3.68	0.42
Instructional Coach F				**Teachers for Instructional Coach F**			
QI	1	3.43	N/A	**QI**	4	3.43	0.42
BM	1	3.67	N/A	**BM**	4	3.53	0.52
RLC	1	3.03	N/A	**RLC**	4	3.61	0.48
Instructional Coach G				**Teachers for Instructional Coach G**			
QI	1	3.64	N/A	**QI**	13	3.64	0.44
BM	1	3.69	N/A	**BM**	13	3.61	0.44
RLC	1	3.76	N/A	**RLC**	13	3.50	0.56
Instructional Coach H				**Teachers for Instructional Coach H**			
QI	1	3.15	N/A	**QI**	20	3.76	0.32
BM	1	3.19	N/A	**BM**	20	3.73	0.36
RLC	1	2.12	N/A	**RLC**	19	3.60	0.49
Instructional Coach I				**Teachers for Instructional Coach I**			
QI	1	2.96	N/A	**QI**	9	3.73	0.34
BM	1	3.24	N/A	**BM**	9	3.55	0.47
RLC	1	2.66	N/A	**RLC**	7	3.64	0.47

Table 4

Principal Components Analysis with varimax rotation and Kaiser normalization

Quality Instruction													
Question	N	Component			Question	N	Component			Question	N	Component		
		1	2	3			1	2	3			1	2	3
GFS # 1	105	.255	***.771***	.353	ISS # 4	100	.480	.449	.498	ES # 12	103	**.691****	.455	.342
GFS # 2	104	.203	***.821***	.390	ISS # 5	102	.437	.345	**.672****	ES # 13	103	*.565**	.329	*.590**
GFS # 3	104	.405	***.806***	.237	ISS # 6	103	.215	.453	***.728****	ES # 14	104	*.620**	.415	.447
GFS # 4	104	.430	***.793***	.231	ISS # 7	103	*.573**	.362	*.571**	ES # 15	104	***.741****	.374	.327
GFS # 5	105	.404	***.804***	.180	ISS # 8	104	.523	.350	*.571**	ES # 16	103	***.756****	.384	.205
GFS # 6	104	.358	***.756***	.241	ISS # 9	100	.486	.422	*.584**	ES # 17	101	***.755****	.461	.294
GFS # 7	103	.328	***.741***	.376	ISS # 10	101	.445	.370	**.679****	ES # 18	103	***.743****	.431	.187
GFS # 8	104	.223	***.769***	.415	ISS # 11	102	.403	.334	***.722****	ES # 19	103	.513	.362	*.563**
GFS # 9	105	.413	**.645***	.398	ISS # 12	98	.470	.329	**.687****	ES # 20	104	***.753****	.316	.360
GFS # 10	104	.456	*.555**	.460	ISS # 13	104	.410	.348	***.712****	ES # 21	104	**.677****	.275	.465
GFS # 11	104	.480	.534	.523	ES # 1	104	*.607**	.427	.533	ES # 22	104	***.783****	.220	.360
GFS # 12	100	.397	**.709***	.384	ES # 2	102	**.649****	.295	.467					
GFS # 13	103	.497	.534	.521	ES # 3	104	**.679****	.315	.412					
GFS # 14	102	.537	*.581**	.378	ES # 4	102	**.641****	.221	.527					
GFS # 15	103	.268	.461	**.649****	ES # 5	103	*.561**	.339	.443					
GFS # 16	103	.421	**.646***	.421	ES # 6	104	***.728****	.398	.373					
GFS # 17	103	.403	*.580**	.466	ES # 7	103	*.616**	.376	.489					
GFS # 18	101	.359	**.644***	.502	ES # 8	103	.540	.462	*.566**					
ISS # 1	104	.436	.519	.500	ES # 9	105	***.737****	.298	.443					
ISS # 2	96	.363	.256	***.774****	ES # 10	102	**.696****	.298	.366					
ISS # 3	95	.328	.348	***.765****	ES # 11	104	***.747****	.396	.297					

Based on Comrey and Lee (1992) interpretation guideline: ***** ***Excellent***; ****** **Very Good**; * *Good*

Table 5

Principal Components Analysis with varimax rotation and Kaiser normalization

Behavior Management														
Question	N	Component			Question	N	Component			Question	N	Component		
		1	2	3			1	2	3			1	2	3
GFS # 1	104	*.616**	***.647****	.314	ISS # 6	95	.346	.469	***.664****	ES # 13	100	****.771********	.353	.373
GFS # 2	100	.465	***.659****	.406	ISS # 7	97	***.705****	.356	.447	ES # 14	101	***.636****	***.673****	.303
GFS # 3	105	.515	*.615**	.344	ISS # 8	97	***.669****	.439	.510	ES # 15	103	***.700****	.458	.440
GFS # 4	105	.371	****.738********	.448	ISS # 9	98	.511	.306	*.612**	ES # 16	99	***.650****	.384	.526
GFS # 5	102	*.565**	.436	.544	ISS # 10	99	***.697****	.392	.509	ES # 17	98	*.592**	.485	.408
GFS # 6	100	***.636****	.494	.476	ISS # 11	99	***.667****	.537	.403	ES # 18	96	*.558**	.483	*.569**
GFS # 7	101	***.693****	.380	.472	ISS # 12	95	.296	.484	****.742********	ES # 19	99	***.687****	.334	.507
GFS # 8	101	*.582**	.482	.525	ISS # 13	99	.437	.501	*.606**					
GFS # 9	91	****.734********	.515	.301	ISS # 14	99	.358	.547	*.585**					
GFS # 10	102	*.586**	***.687****	.262	ES # 1	102	.390	.494	***.650****					
GFS # 11	103	.430	***.702****	.436	ES # 2	99	.445	.485	*.551**					
GFS # 12	100	.508	***.663****	.393	ES # 3	100	.343	.442	****.723********					
GFS # 13	100	.390	****.710********	.420	ES # 4	99	.506	.503	*.579**					
GFS # 14	97	.291	*.623**	*.555**	ES # 5	96	*.568**	.475	.541					
GFS # 15	98	.277	****.766********	.449	ES # 6	100	*.578**	.485	.532					
GFS # 16	102	.408	***.709****	.449	ES # 7	100	*.587**	*.551**	.447					
ISS # 1	98	*.563**	.295	*.610**	ES # 8	103	***.658****	.529	.351					
ISS # 2	102	.527	.371	***.663****	ES # 9	96	***.641****	*.599**	.303					
ISS # 3	93	.468	.241	***.649****	ES # 10	101	***.658****	.340	.510					
ISS # 4	98	***.674****	.387	.533	ES # 11	102	*.586**	.431	.465					
ISS # 5	97	.486	.339	****.713******	ES # 12	99	*.617**	.511	.425					

Based on Comrey and Lee (1992) interpretation guideline: ***** ***Excellent***; ****** **Very Good**; * *Good*

Table 6

Principal Components Analysis with varimax rotation and Kaiser normalization

Responsive Learning Communities														
Question	N	Component			Question	N	Component			Question	N	Component		
		1	2	3			1	2	3			1	2	3
GFS # 1	85	**.636****	.266	*.598**	ISS # 1	89	*.626**	.387	*.587**	ES # 1	91	.380	***.847*****	.156
GFS # 2	83	*.583**	.299	**.660****	ISS # 2	93	**.648****	.534	.447	ES # 2	92	.280	***.835*****	.291
GFS # 3	86	**.704****	.259	.530	ISS # 3	88	.345	.477	***.729*****	ES # 3	86	.350	***.759*****	.329
GFS # 4	96	***.754*****	.320	.464	ISS # 4	88	.354	.433	***.773*****	ES # 4	93	.309	***.796*****	.410
GFS # 5	88	**.682****	.303	*.574**	ISS # 5	86	.357	.492	***.732*****	ES # 5	88	.341	***.775*****	.429
GFS # 6	90	*.568**	.480	.548	ISS # 6	92	***.786*****	.467	.297	ES # 6	85	.525	**.653****	.416
GFS # 7	96	*.581**	.538	.477	ISS # 7	93	***.711*****	.486	.443	ES # 7	91	.516	*.600**	.456
GFS # 8	99	**.704****	*.551**	.264	ISS # 8	85	**.634****	.343	**.636****	ES # 8	88	.514	*.597**	.501
GFS # 9	93	**.683****	.511	.373	ISS # 9	92	***.702*****	.461	.449	ES # 9	86	.332	***.741*****	.487
GFS # 10	96	***.736*****	.503	.349	ISS # 10	90	**.665****	.484	.490	ES # 10	89	.526	*.627**	.445
GFS # 11	92	.511	.420	*.601**	ISS # 11	92	*.612**	.543	.473					
GFS # 12	92	.535	.461	*.567**	ISS # 12	94	*.553**	*.603**	.490					
GFS # 13	89	.521	.397	**.695****	ISS # 13	92	*.598**	*.620**	.381					
GFS # 14	95	.514	.512	.357										
GFS # 15	94	**.677****	.542	.317										

Based on Comrey and Lee (1992) interpretation guideline: ***** ***Excellent***; ****** **Very Good**; * *Good*