ASSESSING TEACHERS' SKILLS IN DETECTING AND IDENTIFYING EFFECTIVE INTERACTIONS IN THE CLASSROOM

Theory and Measurement

ABSTRACT

Contemporary education reforms focus on assessing teachers' performance and developing selection mechanisms for hiring effective teachers. Tools that enable the prediction of teachers' classroom performance promote schools' ability to hire teachers more likely to be successful in the classroom. In addition, these assessment tools can be used for teacher training and preparation that contributes to improved student performance. This article summarizes the theoretical and empirical support for a direct assessment of teachers' skill in detecting and identifying effective classroom interactions—the Video Assessment of Interaction and Learning (VAIL). Findings from a study of 270 preschool teachers suggest that the VAIL reliably measures teachers' interaction detection and identification skills. Teachers who can accurately detect effective interactions on video exemplars tend to have more years of education and display more effective interactions with the students in their classroom. Findings are discussed in terms of the implications for teacher selection, preparation, and training.

Faiza M. Jamil
CLEMSON UNIVERSITY

Terri J. Sabol
NORTHWESTERN UNIVERSITY

Bridget K. Hamre
Robert C. Pianta
UNIVERSITY OF VIRGINIA

EFFECTIVE educators promote students' cognitive and social development, which in turn facilitate lifelong learning and success (McCartney, Dearing, Taylor, & Bub, 2007; Pianta et al., 2009). The importance of effective educators is highlighted by the focus on improving selection, preparation, and evaluation of teachers as part of contemporary education reforms targeting student

performance (e.g., Goe, 2007; Nicholson-Crotty & Staley, 2012). Yet in order for these reform initiatives to be successful, there is a pressing need for rigorous assessment tools that can be used to inform teacher development at important junctures, such as certification, selection, and hiring (Glazerman et al., 2011; MET Project, 2010), as well as for teacher training, preparation, and professional development that directly contribute to improved student performance. The current study focuses on the development and validation of a measure that assesses teachers' skills in detecting and identifying effective interactions in the classroom. This concept of teachers detecting and identifying effective practices is not new to the profession. For decades, teachers have spent teacher preparation and professional development time observing other teachers' classrooms, formally and informally. Yet there have not been many efforts to explicitly define, operationalize, and measure this skill to assess how teachers view classrooms, and promote effective teaching through observation (Star, Lynch, & Perova, 2011). We hypothesize that this is a core skill, the reliable measurement of which can provide useful information for selecting and training teachers.

This article summarizes the theoretical and empirical support for a direct measure of teachers' skills in detecting and identifying effective classroom interactions—the Video Assessment of Interaction and Learning (VAIL). Specifically, the current study examines the extent to which teachers' skills in detecting effective interactions through video case examples can reliably be measured as well as the extent to which this skill is associated with teachers' abilities to enact similar practices in their own teaching. This research has implications for enhancing reform initiatives that focus on selecting, hiring, and preparing effective teachers.

The Role of Detecting and Identifying Effective Teaching Interactions

An expanding body of theoretical support and empirical research suggests that the ability to detect effective interactions in the context of teaching may be an essential precursor to consistently demonstrating these skills in the classroom (Hamre et al., 2012; Koran, Snow, & McDonald, 1971). Moreover, emerging evidence suggests that video case examples may uniquely support teacher learning by providing opportunities for targeted observations of key constructs (e.g., Borko, Jacobs, Eiteljorg, & Pittman, 2008; Hatch & Grossman, 2009; Santagata, Zannoni, & Stigler, 2007).

Although research on observing and identifying interactions in the classroom is somewhat less developed than other elements of teacher performance, the focus on teachers' ability to detect effective interactions is not new. In 1904, John Dewey discussed the importance of teachers acquiring the ability of psychological observation and interpretation. Dewey suggested that teachers should observe other teachers' classrooms with the intention of understanding how teachers and students acted upon and interacted with one another, rather than focusing on teachers' pedagogical skills or skills within a particular content area.

A more clearly explicated theoretical foundation for the role of identifying effective teaching practices in improving teacher effectiveness is based on social learning theory, which posits that individuals learn how to behave largely through observation of others (Bandura, 1986). More specifically, this theory suggests that learning does not occur solely through memorizing decontextualized facts, or trial and error in one's own actions, but through careful observations of others' actions and conse-

quence. Observing the result of a particular course of action can actually change an individual's thinking and interpretation of a situation.

Information processing and dynamic memory theory extend social learning theory by suggesting that observing expert teachers interacting with students in the classroom context may help less experienced teachers to develop schemas, or cognitive networks, and scripts for effective teaching practices (Moskowitz, 2005; Schank, 1997). These scripts are important because they serve as a resource that can be quickly drawn on for an effective course of action in the classroom when a similar situation arises in the future. This efficiency of information processing is posited to help teachers improve their interactions with students in complex classroom contexts, where the cognitive demands of attending to varying needs of many students can interfere with teacher performance (Feldon, 2007). Video observations may especially promote efficient information processing by providing observational opportunities that are rich in content and contextual clues that can repeatedly be watched (Sherin & van Es, 2005). Taken together, both of these theories suggest that observation may be critical for developing teachers' skills in the classroom.

In addition to theoretical support, the skill of detecting and identifying is a well-documented component of expertise in a wide range of disciplines, ranging from medicine to athletics to aviation (Bransford, Brown, & Cocking, 2000; Miller, 2011). Researchers have found that experts are able to better discriminate and interpret meaningful patterns in their areas of expertise (Sherin, Jacobs, & Philipp, 2011). For example, past research has demonstrated that expert and novice teachers see features of classrooms differently (e.g., Kerrins & Cushing, 2001). When viewing classroom video, novice teachers focus more on the teacher instead of the students, and view the classroom as a series of disconnected events, whereas expert teachers are better able to organize the classroom interactions and interpret the instructional strategies and effectiveness of the teacher (Sabers, Cushing, & Berliner, 1991). This efficiency is posited to help teachers improve their moment-to-moment interactions with students (Feldon, 2007). Being able to isolate important information from the environment is an important skill for teachers because classrooms are complex contexts (Downer, Jamil, Maier, & Pianta, 2011). Teachers must be able to attend to important classroom cues and infer what they indicate with respect to multiple children's learning. When deciding how to respond to complex classroom situations, expert teachers appear to chunk information and quickly assess the potential outcomes of various courses of action (Miller, 2011).

In addition, teachers' skills in detecting and identifying effective classroom interactions may be important for teacher preparation and training. For example, Van Es and Sherin (2002) focused on teacher "noticing," which refers to what teachers attend to and how they make meaning of classroom interactions and instructional activities. They conducted a study in which teachers watched a series of classroom videos and discussed what they saw in terms of teachers' roles, student thinking, and classroom discourse. After viewing the videos, teachers shifted from more surface observations to interpreting the events and understanding the effectiveness of the teacher's approach to student learning. Results suggest the utility of videos to support the development of teachers' skills in understanding what is happening in their classrooms (Sherin & van Es, 2005). They also indicate that measuring this skill in a reliable way may be important for teacher training and professional development so that it can be tracked and targeted systematically.

Providing a Lens for Teacher Observations of Effective Teaching

Any systematic analysis of what teachers are observing—for conceptual, research, or professional development purposes—requires a language and lens to label and code what teachers describe based on their observations. Empirically supported theory suggests that interactions between children and teachers are the primary mechanism of development and learning in the classroom (Hamre & Pianta, 2010; Howes et al., 2008). The current study draws upon an existing standardized framework for observing interactions in the classroom, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), to determine if observed interactions can, in fact, be considered effective.

The CLASS is a widely used, validated classroom observation tool that can be used to systemically code the quality of teacher-child interactions (Mashburn et al., 2008). Observations coded with the CLASS have clearly established links to children's learning (e.g., Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Hamre & Pianta, 2005; Wasik & Hindman, 2011). In this study, we focus in particular on interactions related to instruction because research suggests that children do not typically experience high-quality instructional interactions, despite their importance for children's learning (Justice, Mashburn, Hamre, & Pianta, 2008). The VAIL targets three main components of teachers' instructional support—instructional learning formats, quality of feedback, and language and literacy support—all of which have important associations with children's development and learning.

Instructional learning formats focus on the strategies teachers use to promote student interest, engagement, and learning from classroom activities. In classrooms with high-quality instructional learning formats, teachers engage students in active learning by providing interesting and creative materials, orienting children toward the learning objectives of classroom activities (Copple & Bredekamp, 2009; Matheson & Shriver, 2005), and facilitating their involvement in lessons (Penuel et al., 2012). Studies have shown that providing instructional clarity and describing the objectives of learning activities can promote autonomy and self-regulation and increase the time that children are able to engage independently in learning tasks (Cameron, Connor, & Morrison, 2005). Furthermore, exposing children to exciting, creative, and, novel academic experiences can help foster the development of positive learning behaviors (Raab, Dunst, Wilson, & Parkey, 2009). Allowing young children opportunities to participate in a variety of creative learning capitalizes on their early interests and promotes students' engagement in school (Neitzel, Alexander, & Johnson, 2008).

Quality of feedback in a classroom includes interactions through which the teacher expands students' learning and encourages continued participation. Promoting the development of metacognitive skills, or awareness and understanding of one's own thinking processes by asking students to explain their thinking, is an important component of feedback and is important for academic development (Veenman, Kok, & Blöte, 2005). In addition, scaffolding, in which teachers engage in back-and-forth exchanges with students in an effort to deepen students' understanding, has been associated with increased language and social development (Barnett et al., 2008; Justice, Mashburn, Pence, & Wiggins, 2008) and academic performance (Howes et al., 2008).

Language and literacy interactions reflect teachers' skills in stimulating and facilitating language and literacy development, such as engaging in frequent conversations, asking open-ended questions, using advanced vocabulary, and referencing print in the classroom. In classrooms where teachers use a variety of words and connect them to more familiar words, children demonstrate growth in vocabulary, communication, and language use (Bierman et al., 2008). The use of open-ended questions has also been related to gains in vocabulary (van Kleeck, Vander Woude, & Hammett, 2006) and language development (Walsh & Blewitt, 2006), while language modeling in the classroom improves preschoolers' behavioral self-regulation and socioemotional competence (McClelland, Cameron, Wanless, & Murray, 2007).

Unfortunately, research suggests that many children do not consistently experience high-quality instructional interactions (Justice, Mashburn, Hamre, & Pianta, 2008). This unevenness in instructional support in the classroom, coupled with the evidence that instructionally supportive interactions matter for children's success, suggests that measuring teachers' ability to detect these interactions in the classrooms may provide important information regarding teachers' skills and effectiveness (e.g., Pianta et al., 2008). We hypothesize that being able to detect behaviors and skills in this domain plays an important role in developing and executing those behaviors in the classroom.

Measuring Teachers' Skills in Detecting Effective Interactions: The Video Assessment of Interactions and Learning (VAIL)

Given the current need for validated teacher assessments and the growing theoretical and empirical support for teachers' skills in detecting and identifying effective classroom interactions as a target for teacher learning, the development of a measure that assesses this skill is an imperative next step. The newly developed Video Assessment of Interactions and Learning (VAIL) directly assesses teachers' skills in detecting effective classroom interactions from video. The CLASS serves as a guiding framework to organize the wide range of interactions in classrooms that are associated with positive child outcomes (Hamre et al., 2013) that teachers can directly observe using the VAIL.

The VAIL consists of two short videos that highlight specific dimensions of instructional support in the CLASS. Teachers view these videos and respond to two prompts focused on a particular dimension of classroom interactions for each video. Each prompt asks teachers to identify five strategies the teacher in the video is using to facilitate a particular type of effective instructional interaction in the classroom as well as specific, behavioral examples of each strategy. Teachers' skills are then coded based on the ability to identify correct and specific instances of effective interactions of the required dimension and the extent to which the answers represent a breadth of understanding across the dimension (see the Method section for further detail on development and implementation of the VAIL). This approach is unique because the scoring is organized around and informed by a model of effective classroom interactions that has strong empirical support.

Current Study

Although teachers regularly observe in classrooms, in large part there has been a failure to operationalize and measure the skill of identifying effective practices specifically and consistently. We developed the VAIL based on the current theoretical and empirical support that the ability to identify effective skills is a critical link to exhibiting those skills in practice. In the remainder of the article, we summarize results on (1) the extent to which teachers' skill in detecting interactions can be measured reliably and (2) the extent to which this skill relates to their demographic characteristics and effectiveness in the classroom.

The first aim of this article addresses the utility of the VAIL as a psychometrically sound measure of teachers' observation skills. More specifically, we explore the factor structure, internal consistency, and interrater reliability of the VAIL. For the VAIL to be a useful tool in teacher hiring and development decisions, it is essential that the results it provides consistently measure a construct and can be reproduced by multiple raters. We present initial evidence that the scores from the VAIL can be trusted and are free from random error within an acceptable margin (Bollen, 1989a).

Second, this article aims to provide initial evidence on the validity of the VAIL as a measure of teachers' skill in detecting and identifying effective classroom interactions. More specifically, we explore which teacher characteristics are associated with performance on the VAIL and the degree to which teachers' VAIL scores predict the quality of observed teacher-child interactions in their own classrooms. For the VAIL to be useful, it is important to establish that it measures the construct that it is intended to measure. One way to do this is to examine the ways in which it relates to other variables and ensure that these relationships are aligned with accepted theory (Kane, 2006). Based on the extant theoretical literature related to teacher learning and development, we expect teachers' directly assessed skill in detecting effective classroom interactions to be associated with indicators of expertise in teaching, such as years of education and experience, but not with personal demographic factors such as age. We also expect teachers' scores on the VAIL to predict the quality of interactions in their own classroom, especially those interactions that are of an instructional nature.

Method

Participants

Participants in this study included 270 preschool teachers drawn from a larger sample of teachers involved in a randomized controlled trial testing the effects of a 14-week professional development course on effective classroom interactions. The study was conducted in 10 sites across the United States; results from the intervention are described elsewhere (Hamre et al., 2012). The course, entitled Support of Language and Literacy Development in Preschool Classrooms through Effective Teacher-Child Interactions and Relationships, had two main instructional targets: the first focused on effective teacher-child interactions, and the second focused specifically on the teaching of literacy and language in early childhood classrooms.

Recruitment for the main intervention study focused on large community-based and Head Start preschool programs across the country, and enrollment took place in two cohorts; Cohort 1 started the course intervention across five sites in the spring of

2008. Teachers in this cohort were administered the VAIL before and after the intervention. Cohort 2 added five more sites in the spring of 2009. Because of the excessive length of the precourse survey in which the VAIL was embedded, teachers in the second cohort did not complete a precourse VAIL. They only completed the VAIL at the end of the intervention.

Before the start of the study, teachers and program administrators were invited to attend recruitment meetings at each site. Teachers were eligible for the study if they (1) were the lead teacher in a classroom in which the majority of children were eligible for kindergarten the following school year, (2) conducted instruction in English for the majority of the school day, and (3) had access to high-speed internet. The subsample of teachers used for this smaller study includes all teachers who completed a VAIL free of intervention effects ($N = 270$), meaning intervention and control teachers from the first cohort who completed the VAIL prior to the start of the intervention ($N = 191$), and control teachers from the second cohort, who completed the VAIL after the end of the intervention period ($N = 79$). Before combining these two samples, we tested for cohort differences on the VAIL total score and found no statistically significant mean differences between cohorts, $t(224) = 1.29$, $p = .197$.

A total of 427 teachers were recruited for the study (Cohort 1 $N = 249$; Cohort 2 $N = 178$). Overall, 103 (Cohort 1 $N = 74$; Cohort 2 $N = 28$) participants did not participate or left the study due to various reasons including relocation, health issues, and inability or unwillingness to attend course sections (treatment group only, $N = 53$). The analytic sample for this study was drawn by including all teachers who agreed to participate in the study and completed VAIL assessments that were independent of intervention effects ($N = 270$). This included preintervention and control teachers from Cohort 1 who completed the preintervention survey, and control teachers from Cohort 2 who completed the postcourse survey.

The teachers in this sample ($N = 270$) were diverse in terms of personal and professional demographics. The sampled teachers had an average age of 42.74 years ($SD = 10.5$). Most participants were women (93%) and most were Black (43.1%), White (27.7%), or Hispanic (19.9%). Participants had an average of 15.62 years of education ($SD = 1.65$) and had been teaching for an average of 14.16 years ($SD = 9.34$). Of the sampled teachers, 78% held a child development associate's degree (CDA) or teacher certification in pre-K or kindergarten.

Procedures

Data collection. Data for this study were collected at different time points during the study year. Teacher demographic data were collected using an online survey before the start of the intervention window, in the spring of the study year. For teachers in Cohort 1, the VAIL was embedded in the precourse survey. For teachers in Cohort 2, the VAIL was part of a postcourse online survey, administered at the end of the 14-week intervention window. In order to obtain teachers' observation data, all teachers were provided with a digital video camera and digital video (DV) cassettes at the start of the study, with detailed documentation and training on how to use the camera. Teachers sent in four DVs during the study, each recording 30 minutes of class time, from which two 15-minute segments were double-coded using the CLASS by randomly assigned coders. For this study, we use an aggregate score of the first two DVs (four 15-minute segments) of classroom video to establish each classroom in-

teraction score in order to obtain a stable estimate of each classroom's interaction quality. This time frame is not ideal for capturing clean assessments of teachers' effective classroom interactions, as some of the DVs were recorded after some of the teachers had started receiving the intervention. Although we anticipate that the use of this time frame for videos may decrease relations between the VAIL and teachers' observed classroom interactions, recent work has suggested that using more than one video of classroom practices can enhance the reliability of estimates of teachers' use of effective interactions (Mashburn, Downer, Rivers, Brackett, & Martinez, 2014).

Development of the VAIL coding. Creation of the scoring manual for the VAIL was an iterative process during which a team of researchers watched numerous consented videotapes of preschool classrooms from past projects. The researchers selected two video clips that showed multiple occasions of instructionally focused interactions between teachers and children. Once clips were selected, the researchers wrote two prompts per video clip, asking participants to identify effective interaction strategies teachers in the video were using, and specific examples of those strategies from the video. The first prompt for each video focused on the quality of a particular type of instructional interactions, such as instructional learning formats, or the quality of feedback. The second prompt for each video asked teachers specifically about language- and literacy-related interactions because this content area was a target of the course that the intervention teachers were taking as part of the larger study. A researcher trained in the Classroom Assessment Scoring System (CLASS) then transcribed all instances from the video clips that matched instructionally focused interaction types from the CLASS. This document served as the first draft of the coding manual. The team of researchers then used this manual to assign codes to a sample of teacher responses, adjusting the language of the manual to reflect any missed interactions or to provide greater clarity. Following this iterative process, the final product was used to code the VAIL which was part of this study.

VAIL coder training. In order to become reliable VAIL coders, research assistants attended a half-day training session during which they viewed the same video clips that teachers taking the VAIL watched. During training, they also read and discussed the VAIL manual and coded and discussed sample teacher responses. They then independently coded a complete practice VAIL and had an opportunity to ask questions before taking a reliability test. Reliability was established by obtaining an 80% match on two complete video assessments (80 items each) between the research assistants' codes and the master codes determined by three VAIL experts. If research assistants were not reliable on their first attempt, they had additional training and discussion and then completed coding of two more video assessments to establish reliability.

In an effort to maintain ongoing reliability, research assistants were all required to code the same video assessment as a drift test on a weekly basis. Master codes for these drift tests were based on the VAIL trainer's codes, and research assistants had to obtain at least 80% agreement with the master codes in order to be considered reliable. Drift meetings were used to clarify coding problems from the week and practice discussing justifications for the master codes based on the contents of the coding manual.

CLASS coder training. In order to become reliable observers using the CLASS (Pianta et al., 2008), research assistants attended a 2-day CLASS training. This in-

Table 1. Prompts and Corresponding Strategies and Behavioral Examples for the Video Assessment of Interactions and Learning (VAIL)

Prompts	Strategies	Behavioral Examples
Video Clip 1:		
Name up to *5* strategies the teacher is using *to engage the students in the lesson and hold their attention*. For each strategy, list a specific, behavioral example of the strategy from the clip.	Advanced Organizer	Teacher tells children they will be looking for a word in the message on the board.
	Effective Questioning	Teacher asks a child, "Can you find the word 'play'? What letter should you be looking for? What is the first letter in the word 'play'?"
Name up to *5* strategies the teacher is using *to develop children's language and/or early literacy skills*. For each strategy, list a specific, behavioral example of the strategy from the clip.	Self-Talk	The teacher describes her actions as she writes the words on the board.
	Models Writing	The teacher writes P/L/A/Y on the board.
Video Clip 2:		
Name up to *5* specific, observed behaviors that the teacher uses *to effectively provide feedback and extend students' learning, skills, and persistence*. For each strategy, list a specific, behavioral example of the strategy from the clip.	Assistance	When the boy asks the teacher what his word is, she says, "Well let's sound it out. What sound does the 'a' make?"
	Reinforcement	When the student succeeds in finding her word, the teacher says, "That's it! You found it!"
Name up to *5* strategies the teacher is using *to develop children's language and/or early literacy skills*. For each strategy, list a specific, behavioral example of the strategy from the clip.	Sounds Out Words	The teacher sounds out the word "and" with one of the boys.
	Repetition and Extension	When the boy says "I don't have red," the teacher says, "You don't have red? Use this color."

volved reviewing each of the 10 dimensions of the CLASS in detail, and then watching, coding, and discussing multiple 20-minute videos of classroom teaching. They then had to pass the CLASS reliability test, which requires scoring five videos and demonstrating 80% consistency within one point of the master codes. Coders were given one additional chance to achieve reliability before being assigned other tasks on the project. In order to maintain reliability, coding meetings occurred weekly and all coding team members coded the same video segment, randomly selected from actual project data. Coders were required to maintain 80% reliability in order to continue coding for the project.

Measures

Video Assessment of Interactions and Learning (VAIL). As part of the precourse and postcourse online survey, teachers watched two teaching video clips online and responded to a prompt asking them to provide particular types of strategies and examples. The wording of all four prompts and a sample of acceptable strategy and example responses can be found in Table 1. The first video shows whole-class instruction, in which children are sitting on the carpet and reading a message that the teacher has written on a chart. The teacher reads with them and uses a sheet of yellow plastic to track the works as they read. She then asks individual students to come up and circle a particular letter or find certain words in the message. For this video, participants identified five strategies and five corresponding behavioral examples that reflected

interactions the teacher was using to encourage student engagement (Instructional Learning Format Prompt) and five strategy-example pairs that exemplified interactions the teacher was using to promote children's language and literacy development (Language and Literacy Prompt 1).

Examination of the teacher response to the first prompt in Table 1 shows that the teacher has identified the use of creative materials as the strategy, and identified a specific example from the video when the teacher was using a creative material to keep students engaged. This response would then receive four separate codes (described below), as would each of the other four teacher responses for instructional learning formats and the five strategy-example pairs the teacher wrote for language and literacy.

For the second video, participants identified five strategy-example pairs related to the feedback the teacher provided to students during a learning activity (Quality of Feedback Prompt) and another five strategy-example pairs related to children's language and literacy development (Language and Literacy Prompt 2). This video, entitled *Letter Hunt*, shows students working individual assignments at their desks. They have a piece of writing in front of them, and they have a post-it note with a letter on it that they have to find in the writing. The teacher walks around the classroom, observing students and offering assistance as needed. She also incorporates instruction on the different sounds the letters make, and the difference between upper- and lower-case letters. The following is a sample teacher response to the language and literacy prompt for this video:

> Strategy: They develop an awareness of the alphabet through "playing" with words.
> Example: The teacher encourages the children to sound out words and circle all of the p's in blue.

Each strategy-example pair was coded for accuracy based on four criteria: (1) strategy: accurate indicator of the interaction being targeted in the particular video prompt; (2) example: specific, behavioral example from the video; (3) match: behavioral example matched strategy identified; and (4) breadth: the degree to which the participants' responses ranged across the different indicators for a type of interaction. Strategy, example, and match codes are dichotomous (1 or 0) and represent whether the strategy, example, or match provided by the teacher is correct or incorrect. Breadth scores are continuous and can range from 0 (if a teacher identified no correct strategies) to 5 (if each strategy identified by a teacher came from a unique category of possible correct answers). These four codes are discussed in more detail below.

Strategy. The strategy is correct if it matches one of the types of behaviors from the CLASS listed in the answer key in the coding manual. This code helps establish whether teachers understand and can articulate which types of interactions are deemed effective according to the CLASS. For teachers to detect and identify effective interactions, they have to have a cognitive organizational structure, or schema, for them. They also must have the language with which to describe these interactions (Sweller, 2005). Based on the VAIL manual, in the literacy-related sample response above, the teacher identified the strategy being portrayed in the video as the development of alphabet knowledge, so this strategy would be scored as correct.

Example. An example is correct if it matches any of the instructionally focused interaction instances listed in the answer key from the video. This code helps determine whether teachers are able to apply their schema for effective classroom interactions to what they are observing—meaning that their schema is organized enough to guide their attention so that they are able to pick out only the specific behaviors that are effective according to the CLASS lens (Downer et al., 2011). In the literacy example above, the participant has identified the sounding out of words as the specific example of what the teacher is doing to develop children's alphabet knowledge, so this would be coded as a correct example.

Match. A strategy-example match is correct when the example presented from the video is not only an effective, instructionally focused behavior listed in the answer key, but is also an accurate example of the particular type of strategy described. This code ensures that the teacher's schema for effective interactions can be applied with the greater specificity and depth of knowledge that comes with expertise (Feldon, 2007). In the literacy example above, the teacher would not receive a correct score for match because even though she identified a correct strategy (developing alphabet knowledge) and a correct example (identifying the sounds letters make) that support language and literacy development in the video clip, the example she provides is actually one that helps develop children's phonological awareness, not their alphabet knowledge.

Breadth. The breadth score measures the degree in which the teacher's responses ranged across the different types of possible strategies present in the video. Since expert understanding of a topic, such as effective interactions, demonstrates both depth and breadth, this code captures the range of interactions that teachers can accurately detect and identify. The VAIL manual organizes strategies into broader categories, and the breadth score indicates which of the categories the teacher has drawn from in her response. The total breadth score is calculated by adding up the number of unique categories represented across the five responses for each prompt.

Whereas alphas for the 20 raw scores assigned to each type of VAIL code (strategy, example, match, and breadth) ranged from .58 to .72, the internal consistency of the overall VAIL score, across all 80 raw codes, for the current sample was .90. Mean and standard deviation for the measure is reported in Table 2.

Observed classroom interaction quality. Interaction quality in participants' classrooms was measured using the Classroom Assessment Scoring System (Pianta et al., 2008). The CLASS is a validated classroom observation tool that can be used to code the quality of teacher-child interactions across 10 distinct dimensions (Mashburn et al., 2008). Previous research demonstrates that these dimensions are organized into three broad domains (Hamre et al., 2012). Positive climate, negative climate (in which a low score is desirable), teacher sensitivity, and regard for student perspectives all fall into the broad domain of Emotional Support. Behavior management, productivity, and instructional learning formats are all components of the Classroom Organization domain. Concept development, quality of feedback, and language modeling comprise the Instructional Support domain. Each dimension of the CLASS is scored on a 7-point scale, with 1–2 representing low scores, 3–5 representing moderate scores, and 6–7 representing high scores. Domain-level alphas for the study sample ranged from .65 to .86. Domain-level means and standard deviations are reported in Table 2.

Table 2. Descriptive Statistics and Intercorrelations among VAIL Factor Scores, Classroom Interaction Domains, and Select Covariates ($N = 270$)

	M	SD	Range	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. VAIL overall	0	.45	2.11	—	.96***	.94***	.94***	.95***	.96***	.94***	.06	.17*	.03	.09	−.01	.17**	−.07	.03
2. VAIL skill	0	.44	2.12		—	.82***	.82***	.99***	.86***	.81***	.05	.16*	.05	.15	.02	.12	−.05	.06
3. VAIL knowledge	0	.42	1.78			—	.99***	.79***	.99***	.99***	.07	.16*	.01	.05	−.02	.20**	−.09	.01
4. VAIL strategy	0	.43	1.83				—	.78***	.97***	.99***	.07	.16*	.01	.05	−.03	.19**	−.08	.01
5. VAIL example	0	.45	2.07					—	.85***	.78***	.04	.16*	.04	.14	.03	.13	−.06	.07
6. VAIL match	0	.32	1.40						—	.98***	.06	.16*	.01	.06	−.01	.21**	−.09	.02
7. VAIL breadth	0	.43	1.82							—	.07	.16*	.01	.05	−.01	.20**	−.09	.01
8. CLASS ES	5.24	.46	2.54								—	.48***	.67***	−.08	.01	.01	.03	−.03
9. CLASS IS	2.54	.60	4.17									—	.52***	.03	−.01	.12	−.02	.07
10. CLASS CO	5.24	.62	3.78										—	−.03	.02	.05	.02	−.09
11. Age	42.7	10.5	49											—	.57***	−.01	−.01	.04
12. Yrs. experience	14.2	9.3	47												—	−.03	−.05	.02
13. Yrs. education	15.6	1.7	8													—	−.06	−.08
14. ECE courses	1.7	1.7	4														—	.04
15. Certified	.78																	—

Note.—CLASS ES = classroom emotional support; CLASS IS = classroom instructional support; CLASS CO = classroom organization; Yrs. experience = the number of years of teaching experience; ECE courses = early childhood education courses.

*$p < .05$.
**$p < .01$.
***$p < .001$.

Teacher characteristics. Participants reported information on their personal demographic characteristics, such as age and ethnicity, at the start of the study. They also reported on professional characteristics such as education level, years of experience, and certification status.

Analytic Plan

Research aims for this study included examining the extent to which teachers' skill in detecting and identifying effective classroom interactions could reliably be measured and understanding how it related to teachers' demographic characteristics and effectiveness in the classroom. Before we could investigate the psychometric properties of the VAIL we needed to create aggregated scores from the raw coded data. For each of the two videos used in the VAIL, there are two prompts, and teachers give five responses per prompt (total = 20 teacher responses). To each of these responses, coders give four codes (strategy, example, match, and breadth), so that each teacher VAIL received 80 raw codes.

In order to reduce the number of parameters needing to be estimated and limit problems associated with the nonnormal distribution of dichotomous data (Bandalos, 2002), for each of the four VAIL prompts, sum score parcels were created across all five teacher responses per prompt for the number of correct strategies, examples, and matches. Along with the total breadth score for each prompt, this resulted in 16 VAIL scores—four for each prompt. To account for scaling differences between the different types of scores, these 16 scores were then standardized by transforming them into z-scores ($M = 0$, $SD = 1$ for all scores). The 16 standardized strategy, example, match, and breadth scores, one of each type for each prompt, were then used to determine the factor structure of the VAIL.

Factor structure and reliability. Confirmatory factor analysis (CFA) was conducted to evaluate whether the 16 composite scores of the VAIL—one strategy, example, match, and breadth score for each prompt— could be combined into latent factors that represented teachers' skill in detecting interactions from video. Competing factor structures were tested, and fit indices, conceptual clarity, and predictive power were used to select the final factor structure. We initially tested a four-factor model with each factor representing one of the four score types and a one-factor structure where all 16 VAIL indicators loaded onto a single overall factor. We hypothesized, based on our conceptualization of teacher skill in detecting and identifying effective interactions as a single, unified construct, that four different types of VAIL codes (strategy, example, match, and breadth) measured using four different prompts (1 for instructional learning formats, 1 for quality of feedback, and 2 for language and literacy) would be best represented by one overall factor. The scaling of the latent factors was established by setting one path in each factor to one. All models in this study were specified using maximum-likelihood estimation on covariance matrices in Mplus 7.0 (Muthén & Muthén, 2012).

In order to assess reliability of the VAIL, we estimated the internal consistency of VAIL scores using Cronbach's alpha (1951). We calculated interrater reliability though the use of kappa, a chance-adjusted measure of rater agreement. Kappa is a more rigorous measure of agreement between two coders than percentage of absolute agreement because it corrects for the possibility that some of the time coders agree solely based on chance (Cohen, 1960).

Associations with teacher characteristics and practices. Once we had arrived at CFA models with good fit, we examined the correlation of VAIL factor scores with teacher professional and personal demographic characteristics. Demographic variables that were correlated with the VAIL were used as covariates in subsequent analyses. In order to understand the association between the VAIL and teachers' classroom interactions, we reran our final CFA model by including significant covariates and all three domains of the CLASS as outcomes, being predicted by the VAIL factor score. In addition, this model in which the outcome variables represented observed interaction quality also controlled for intervention condition. This was done to account for potential intervention effects on CLASS scores, estimates of which were obtained by aggregating over multiple observations through the intervention period, in the interest of reliability.

Results

Factor Structure and Reliability

One-factor and four-factor CFA models were used to empirically test the extent to which the factor structure of our VAIL data fit the hypothesized one-factor model. Both initial models showed extremely poor fit, so we examined the correlation matrix between indicators to inform the next set of models that could be tested. Results indicated moderate to high correlations between indicators within the same prompt (Instructional Learning Format Prompt, range = 0.31–0.95; Language and Literacy Prompt 1, range = 0.53–0.90; Quality of Feedback Prompt, range = 0.21–0.93; Language and Literacy Prompt 2, range = 0.38–0.89). Correlational results also suggested small to moderate correlations by type of code independent of the prompt (strategy, range = 0.31–0.56; example, range = 0.28–0.46; match, range 0.11–0.38; breadth, range = 0.27–0.57). Based on the recommendations by Cole, Ciesla, and Steiger (2007), we tried to fit a multitrait, multimethod model to account for the trait factors of interest (strategy, example, match, and breadth; or overall VAIL factor) as well as the method factors created by the four VAIL prompts. This model did not converge, as is often the case, so we used a correlated uniqueness model to represent the method factors by prompt (Marsh & Bailey, 1991). In this type of model, error terms are correlated, but only between indicators that shared a method (in this case a prompt), in order to account for shared method variance (Cole et al., 2007). We also attempted to fit a model with only the four method factors (four prompts) represented and this model showed very poor fit, increasing our confidence that the good fit of the correlated uniqueness models pointed toward nonignorable, but nonmeaningful, residual relations among methods that needed to be accounted for in our CFA models.

The four- and one-factor models resulted in good fit and acceptable indicator loadings on factor. All indicators loaded onto their respective latent factors loadings greater than 0.3. An examination of the correlations between the factors of the four-factor model showed they were very high (all $r > .95$) between the strategy, match, and breadth factors, suggesting that they represented the same latent construct. Therefore, we also tested a two-factor model in which all indicators for strategy, match, and breadth loaded onto a factor conceptualized as representing *interaction knowledge*, and all the indicators for example were loaded onto a factor conceptualized to represent *skill in detecting interactions*. Correlations between the two factors of this model were also high ($r = .82$).

A range of model fit indices was used to assess the extent to which the data fit the hypothesized model. For a well-specified model, root mean square error of approximation (RMSEA) values and standardized root mean square residual (SRMR) values of .06 or less are an established benchmark reflecting good fit (Hu & Bentler, 1999). The comparative fit index (CFI) and the Tucker-Lewis index (TLI) were also used to examine model fit. Values greater than .90 are typically considered adequate for the CFI and TLI fit indices, but values above .95 are preferable and generally considered to be indicators of good fit (Bollen, 1989b; Bryne, 2001; Hu & Bentler, 1999). Taken as a whole, these statistics indicated that all three models fit the data well, although high correlations between factors in the four- and two-factor models suggested that the more parsimonious one-factor model might be most useful. Multicollinearity between the factors of the four- and two-factor VAIL solutions would make it challenging to include them simultaneously in subsequent predictive models (Cohen, Cohen, West, & Aiken, 2003). Factor loadings, residual correlations, and model fit statistics for the one-factor CFA model with correlated residuals representing method factors are presented in Table 3.

Internal consistency of the overall VAIL measured across all 80 raw scores was high ($\alpha = .90$). When examined separately, the 20 item factors represented by strategy, example, and match scores also came together well, except for match scores, which only had an alpha of .58. These findings provide support for using a one-factor composite for the VAIL.

An interrater reliability analysis using the kappa statistic was performed on the raw VAIL scores to determine consistency among raters. The kappa was 0.57 ($p < .001$) across the 80 items of the VAIL. Kappas ranged from 0.43 to 0.54 ($p < .001$) across the 20 item factors created by strategy, example, match, and breadth scores. These estimates are considered to reflect moderate levels of agreement between raters (Landis & Koch, 1977).

Associations with Teacher Characteristics and Practices

To better understand which personal and professional characteristics of teachers were significantly related to their skill in detecting and identifying effective classroom interactions, we examined the zero-order correlations presented in Table 2. Teachers' years of education was the only demographic variable that was significantly related to teachers' detection skills. The positive correlation, albeit small in size ($r < .21$), suggests that teachers with more education display stronger skills in identifying effective classroom interactions. Although the overall VAIL factor and the factors theorized to be more representative of teacher knowledge of interactions (i.e., strategy, match, and breadth) were significantly related to teachers' years of education, the example score, thought to represent their skill in detecting interactions, was not.

In order to better understand the relation between teachers' skill in detecting and identifying effective classroom interactions and the observed interaction quality in their own classroom, we added in all three CLASS domain scores as outcomes in the final CFA model and predicted them from the VAIL factor score. We hypothesized that teachers' VAIL score, which was based on instructionally oriented video clips, would be significantly associated with the quality of observed instructional interactions in teachers' classrooms, but not with the quality of emotionally and organizationally supportive interactions. This model controlled for teachers' years of educa-

Table 3. Factor Loadings, Residual Correlations, and Model Fit for VAIL One-Factor CFA Model with Correlated Residuals Representing Method Factors ($N = 270$)

| | Standardized Factor Loadings β (SE) | Unstandardized Factor Loadings β (SE) | Residual Correlations | | |
			Strategy	Example	Match
VAIL indicators:					
Prompt: ILF:					
Strategy_ILF	.51 (.06)	1.00(.00)	–		
Example_ILF	.49(.06)	.97 (.18)	.05	–	
Match_ILF	.40(.07)	.79 (.10)	.54***	.17**	–
Breadth_ILF	.51 (.06)	1.14 (.06)	.67***	.02	.50***
Prompt: LL1:					
Strategy_LL1	.66(.05)	1.31 (.22)	–		
Example _LL1	.63(.05)	1.25 (.21)	.11*	–	
Match_LL1	.54(.06)	1.06(.20)	.37***	.32***	–
Breadth _LL1	.62(.06)	1.22 (.21)	.49***	.09	.29***
Prompt: QF:					
Strategy_QF	.53 (.06)	1.05(.19)	–		
Example_QF	.55(.06)	1.08(.20)	−.07	–	
Match_QF	.46(.06)	.91 (.19)	.47***	.20**	–
Breadth_QF	.53(.06)	1.05(.19)	.64***	−.08	.41***
Prompt: LL2:					
Strategy_LL2	.64(.06)	1.27 (.22)	–		
Example_LL2	.61(.06)	1.20(.21)	.10	–	
Match _LL2	.52(.06)	1.03(.20)	.41***	.37***	–
Breadth_LL2	.62(.06)	1.22 (.22)	.49***	.01	.29***
Model fit:					
CFI	.97				
TLI	.96				
RMSEA	.06				
SRMR	.05				

Note.—All factor loadings are statistically significant at $p < .001$. CFI (comparative fit index); TLI (Tucker Lewis index); RMSEA (root mean square error of approximation); SRMR (standardized root mean square residual). ILF = instructional learning formats prompt, LL1 = language and literacy prompt 1, QF = quality of feedback prompt, LL2 = language and literacy prompt 2.

*$p < .05$.
**$p < .01$.
***$p < .001$.

tion and teachers' treatment condition in the larger intervention study. The results for the predictive model are presented in Table 4. Since CFA results were reported in the previous table, and they did not change significantly, this table includes only the results from the prediction model. As hypothesized, results suggest that VAIL scores using the one-factor solution are associated with interaction quality, but only for teachers' instructionally supportive interactions ($\beta = .17$, $p = .02$). Teachers' skills in detecting and identifying effective classroom interactions, as measured by their performance on the VAIL, do not appear to be related to the quality of their emotionally and organizationally supportive classroom interactions.

Discussion

The present study investigated the factor structure and reliability of a new measure of teachers' skill in detecting and identifying effective classroom interactions, the Video

Table 4. VAIL One-Factor Model Predicting Classroom Interaction Quality ($N = 270$)

CLASS Domains	Emotional Support		Classroom Organization		Instructional Support	
	β	(SE)	β	(SE)	β	(SE)
Covariates:						
Condition	.11	(.06)	.06	(.06)	.28*	(.06)
Years of education	.01	(.07)	.05	(.07)	.11	(.06)
VAIL	.07	(.08)	.02	(.08)	.17*	(.07)
Model R^2	.02	(.02)	.01	(.01)	.12**	(.04)
Model fit:						
CFI	.97					
TLI	.96					
RMSEA	.05					
SRMR	.06					

Note.—CFI (comparative fit index); TLI (Tucker Lewis index); RMSEA (root mean square error of approximation); SRMR (standardized root mean square residual).

*$p < .05$.
**$p < .01$.

Assessment of Interactions and Learning (VAIL). It further examined demographic correlates of this skill and the extent to which this skill is related to the observed interaction quality in teachers' own classrooms. This study had several interesting findings. First, the VAIL has a one-factor latent structure, suggesting that it is measuring a single, unified skill. Second, this skill of detecting and identifying effective classroom interactions can be measured with an adequate level of reliability. Third, though not associated with fixed personal demographics of teachers, this skill is related to teachers' years of education. Finally, teachers' performance on this measure is associated with the quality of their classroom interactions in the area of instructional support. These findings suggest that this new assessment can reliably measure a teacher skill that is associated with their classroom instructional practices.

Advancing the Theory and Measurement of Detecting and Identifying Interactions

Even though the VAIL's multiple indicators (i.e., strategy, example, match, breadth) suggest that teachers' skill in detecting and identifying effective interactions may have multiple components, the good fit of a one-factor model suggests that we are measuring a unitary construct. More conceptually, this means that even though we may speak of the skill in terms of two parts, detection and identification, to offer a more complete description, the acts of detecting and identifying are so closely intertwined that it is difficult to measure them separately. In a given moment, for teachers to detect effective interactions they must be able to identify them, meaning they must be able to distinguish them from other interactions that are ineffective, and for teachers to identify interactions as effective they must be aware of them.

This finding also lends credence to the use of information-processing theory in conceptualizing this teacher skill. Moskowitz (2005) described schemas as cognitive associative networks that guide how new information is processed and dictate which information is retrieved from memory. This suggests that what we attend to, and

how we interpret it based on our prior knowledge and experiences, is guided by our schema. In other words, the quality of a teacher's schema for effective interactions will determine their skill in detecting and identifying those interactions. Since the indicators of the VAIL directly assess different qualities of that schema (e.g., specificity and breadth), it makes sense for them to be measured by a single score and represent the construct of teachers' skill in detecting and identifying effective classroom interactions. Furthermore, since the VAIL is based on a validated, standardized language and lens for effective interactions, which has been shown in past research to relate to positive student outcomes (Pianta et al., 2008), teachers' score on this assessment is meaningful.

In addition to having structural qualities that align with its conceptual underpinnings, the validity of a new measure is enhanced when it relates to other constructs in theoretically plausible ways (Kane, 2006). Teachers who displayed greater skill in detecting and identifying effective interactions from video were not demographically different from those who did not, but these teachers did on average have more years of education. These results are promising because they indicate that as an assessment of skill, the VAIL performs equally well for teachers of different age and ethnicity. To the extent that additional years of education provide teachers with more opportunities to learn about effective teaching (Darling-Hammond, 2010), and possibly develop greater skill at identifying it, association between teachers' VAIL scores and years of education is also expected. Interpreted in terms of schema theory, more years of education may offer an opportunity for teachers to develop a more elaborate schema for effective classroom interactions, which enables them to detect and identify these interactions more accurately.

The lack of association between teachers' skill in identifying effective interactions and their years of teaching experience, on the other hand, is unexpected. Because teachers with more experience have had many more opportunities to observe interactions in classrooms and witness the outcomes of these interactions, we would expect more experienced teachers to have stronger skills in detecting and identifying effective interactions. There are two possible explanations for this finding. The first may have to do with the nonlinear trajectory of teacher expertise development. Recent studies have suggested that teachers experience a sharp rise in effectiveness in the first year or two of their careers, but this growth then levels out (Staiger & Rockoff, 2010). It is possible that teachers with considerably more experience are not significantly more skilled at detecting and identifying effective classroom interactions because they developed this skill earlier in their careers and it has since remained stable. Another explanation is that the definition of effective interactions, and the corresponding focus of classroom observations, has changed considerably in the last decade (see Pianta & Hamre, 2009). The VAIL defines effective classroom interactions according to one particular standardized lens for observing classrooms, and it is possible that gains in detection and identification skills that teachers make after a number of years as a result of experience in the classroom are balanced out by a growing misalignment between current conceptualizations of effective interactions and those favored at the time of their initial teacher preparation.

Perhaps the most meaningful finding of this study concerns the association of teachers' skill in detecting effective interactions and the quality of the interactions observed in their classrooms. Based on theories of social cognition, we hypothesized that teachers who could identify effective interactions from video would also be more

effective at interacting with students during the course of their own teaching. Teachers who were more skilled at detecting effective interactions, as measured by their performance on the VAIL, were more instructionally supportive in their interactions with students, but their emotionally and organizationally supportive interactions were not significantly different from teachers with weaker observation skills. Because the VAIL prompts specifically primed teachers to focus on more instructionally oriented interactions—such as those concerning the use of engaging learning formats, effective feedback strategies, and rich language—clear alignment with teachers' own interactions in those areas suggests convergent validity with the empirically validated CLASS observation measure.

Furthermore, the lack of association with teachers' observed interaction in the domains of emotional support and classroom organization provides some evidence for discriminant validity, in that the VAIL was not associated with teacher practice in domains where there was no content alignment. In other words, classroom interactions fall into different domains (e.g., emotional, organizational, instructional), and it is reasonable to expect that skill in detecting effective interactions in one domain would align with the quality of a teacher's interactions in that particular domain. To truly confirm this pattern of validity, we need to develop more modules of the VAIL with prompts priming teachers to detect and identify more emotionally and organizationally supportive classroom interactions. With the current measure, we see a clear alignment in teachers' instructional support detection and interaction skills, but we cannot be sure that this will be the case in other domains of classroom interaction, or that cross-domain associations will not exist.

Taken together, these findings are meaningful because they provide validity and reliability evidence for a theoretically informed measure of a teacher skill related to teachers' classroom behavior. At a time when there is growing need for rigorous assessment tools that can inform teacher hiring and development (Glazerman et al., 2011; MET Project, 2010), these findings have implications for educational reform efforts that need consideration.

Implications for Educational Reform Efforts

In a recent study examining the personnel practices of effective schools, Loeb, Kalogrides, and Béteille (2012) suggested three mechanisms by which schools can control the quality of their teaching force: (1) retention of high-quality teachers, (2) development of existing teachers at the school, and (3) recruitment of high-quality teachers. Each of these mechanisms presents its own unique set of challenges. The loss of effective teachers to better schools, and tenure rules that make it difficult to remove ineffective teachers, make it very difficult to gain traction through the first of these mechanisms (Hanushek, 2009). Considerable funding and research in teacher professional development have led to more consistent, if modest, effects through the second mechanism (Guskey & Yoon, 2009). Yet the area of improved hiring practices is one where the field of education has yet to make considerable inroads.

Due to the dearth of teacher-assessment tools, many school districts have turned to commercially available teacher-selection instruments. A recent synthesis by Metzger and Wu (2008) of 24 studies examining the usefulness of one widely used measure of teachers' affective orientations concluded that the tool had greater alignment with teachers' work ethic than actual teaching ability. Rockoff, Jacob, Kane,

and Staiger (2011) concluded that even though one instrument could not provide enough information to predict teachers' future performance, the use of a broad set of cognitive and noncognitive information may help school districts in selecting more effective teachers. In such a context, an instrument like the VAIL, which can be reliably scored and provides some indication of teachers' actual classroom behaviors, can be a useful selection tool, especially as part of a larger testing battery. The alignment of this instrument specifically with teachers' instructionally supportive interactions, an area where many elementary school students do not experience consistent, high-quality teaching (Pianta, Belsky, Houts, Morrison, 2007), also suggests that the use of this measure has the potential to contribute new and meaningful information to the teacher hiring process.

Beyond measurement, a greater focus on teachers' skill in detecting and identifying effective classroom interactions in professional development and teacher preparation settings also has implications for improving teacher quality. Past research has shown that a precursor to using effective instructional strategies is the ability to recognize those effective strategies in other teachers, and that watching classroom video may be a useful tool in learning to identify these desired behaviors (Pianta & Hamre, 2009; van Es & Sherin, 2002, 2006). A recent study using the VAIL in a sample of preservice teachers found variation in teachers' skill in identifying effective interactions based on how far along they were in their teacher-preparation program (Wiens, Hessberg, LoCasale-Crouch, & DeCoster, 2013). Though this study is cross-sectional, and longitudinal studies are needed to provide stronger evidence, it does suggest that as preservice teachers are learning how to teach, their skills in detecting and identifying effective teaching may also be growing. If longitudinal analyses find that the VAIL has good test-retest reliability, it can potentially be used to measure the growing skill of preservice teachers as they progress through teacher-preparation programs.

Engaging teachers in watching video in an effort to improve their interactions in the classroom is also a component in many coaching interventions for teachers (e.g., Baecher, Rorimer, & Smith 2012; Mashburn et al., 2010; Stephenson, Carter, & Arthur-Kelly, 2011). Yet many coaching-based professional development interventions focus heavily on having teachers watch their own classroom interactions. The extent to which teachers' skill in detecting and identifying interactions in their own classrooms differs from their skill in observing interactions in the classrooms of other teachers is unclear, but there is some indication that these skills may be independent of each other (Seidel, Sturmer, Blomberg, Kobarg, & Schwindt, 2011; Zhang, Lundeberg, Koehler, & Eberhardt, 2011). Further investigation is needed to determine the extent of these differences and the ways in which they might relate to teachers' interactions in the classroom. Considering the logistical intricacies of reliably assessing teachers on what they detect and identify in videos of their individual teaching, the VAIL is a good first step. We have confidence that even teachers' skill in detecting effective interactions, as it is measured by the VAIL, is important in professional development contexts. In the course intervention from which the current study sample was drawn, teachers' skill in detecting and identifying interactions was a significant mediator of intervention effects on effective interactions in teachers' own classroom (Hamre et al., 2012).

The findings from this study, and previous studies of teachers' skills in making observations from video, merit two conclusions. First, teachers' skill in detecting and

identifying effective classroom interactions may be important to consider in contexts of preparation, hiring, and professional development, where teacher expertise is of concern. Second, even though more measurement work is required, there is evidence that this skill can be measured using the VAIL with reasonable confidence in its reliability and validity.

Limitations

We recognize a number of limitations in this initial round of measurement work. One is that observational data of teachers' classroom interaction quality, which served as an outcome of this study, were not entirely free of intervention effects. In order to achieve a more stable estimate of interaction quality, we aggregated across four video segments (Mashburn et al., 2014), some of which were collected after the start of the intervention. The timing of this data collection likely reduced the relation between the VAIL and classroom interaction quality, making our estimates more conservative, but our analyses do not explicitly model the variability in the outcome that results from different videos, time of year, or raters.

Even though initial results pertaining to the factor structure are promising and support the presented conceptualization of detecting and identifying effective inter-actions as a unidimensional skill, it has not been possible to establish clear criterion validity due to a lack of comparable validated measures. Being able to validate the VAIL against another measure of detecting and identifying effective interactions would allow for increased confidence that the VAIL is actually measuring what teachers observe. For instance, the measure potentially captures a construct that is corre-lated with teachers' knowledge of interactions, since this is a requirement for achiev-ing a correct strategy score. Inasmuch as the VAIL is a promising new measure of a teacher skill that is related to effectiveness, and knowledge is another component of teacher effectiveness (Shulman, 1986, 1987), significant overlap in their measurement could be problematic.

A recently published video-based measure by Kersting, Givvin, Sotelo, and Stigler (2010) follows a similar process as the VAIL but is conceptualized as a measure of teacher knowledge. Blomberg, Stürmer, and Seidel (2011) have also validated a video-based assessment of teachers' professional vision, which they define as teachers' ability to observe and make sense of what is happening in a classroom. This construct sounds conceptually closer to the VAIL's construct of detecting and identifying ef-fective interactions, but it is also considerably dependent on teacher knowledge. Further investigation including both of these measures in a study with the VAIL may be a potential next step in this research that would contribute to the theory and measurement of teachers' skill in detecting and identifying effective interactions.

Conclusion

This article summarized the theoretical and empirical support for a video-based, direct assessment of teachers' skill in detecting and identifying effective classroom interactions. The psychometric findings presented here are important because they help establish confidence in the measure itself, but also contribute to our theoretical understanding of the importance of this teacher skill. The findings from this study suggest that the VAIL is associated with the quality of instructionally supportive

interactions that are observed in teachers' classrooms, putting it forth as an instrument that could inform teacher selection and hiring. Furthermore, there is evidence from past research that the VAIL measures a skill that can be improved with practice, and is at the core of several successful professional development interventions for teachers.

Note

References

Baecher, L., Rorimer, S., & Smith, L. (2012). Video-mediated teacher collaborative inquiry: Focus on English language learners. *High School Journal*, **95**(3), 49–61. doi:10.1353/hsj.2012.0007

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, **9**(1), 78–102. doi:10.1207/S15328007SEM0901_5

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River, NJ: Prentice Hall.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, **23**(3), 299–313. doi:10.1016/j.ecresq.2008.03.001

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, **79**(6), 1802–1817. doi:10.1111/j.1467-8624.2008.01227.x

Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education*, **27**(7), 1131–1140. doi:10.1016/j.tate.2011.04.008

Bollen, K. A. (1989a). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1989b). A new incremental fit index for general structural models. *Sociological Methods & Research*, **17**(3), 303–316. doi:10.1177/0049124189017003004

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, **24**(2), 417–436. doi:10.1016/j.tate.2006.11.012

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bryne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, **25**(2), 166–176. doi:10.1016/j.ecresq.2009.10.004

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*, **43**(1), 61–85. doi:10.1016/j.jsp.2004.12.002

Cohen, J. (1960). A coefficient for agreement for nominal scales. *Education and Psychological Measurement*, **20**, 37–46. doi:10.1177/001316446002000104

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, **12**(4), 381–398. doi:10.1037/1082-989X.12.4.381

Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice*. Washington, DC: National Association for the Education of Young Children.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297–335. doi:10.1007/BF02310555

Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, **61**(1–2), 35–47. doi:10.1177/0022487109348024

Dewey, J. (1904). The relation of theory to practice in education. *The third NSSE yearbook (part 1)*. Chicago: University of Chicago Press.

Downer, J., Jamil, F., Maier, M., & Pianta, R. (2011). Using video-based coursework and coaching to improve teacher-child interactions: The role of basic psychological processes. In C. Howes, R. Pianta, & B. Hamre (Eds.), *Effective professional development in early childhood education*. Baltimore: Brookes.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, **42**(3), 123–137. doi:10.1080/00461520701416173

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., Whitehurst, G. J., & Croft, M. (2011). *Passing muster: Evaluating teacher evaluation systems*. Brown Center on Education Policy at Brookings (No. 6971), Mathematica Policy Research, Washington, DC.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development. *Phi Delta Kappan*, **90**, 495–500.

Hamre, B., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, **76**(5), 949–967. doi:10.1111/j.1467-8624.2005.00889.x

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization and measurement. In J. Meece & J. Eccles (Eds.). *Handbook of research on schools, schooling, and human development* (pp. 25–41). New York: Routledge.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . & Scott-Little, C. (2012). A course on effective teacher-child interaction effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, **49**(1), 88–123. doi:10.3102/0002831211434596

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal*, **113**, 461–487. doi:10.1086/669616

Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (pp. 165–80). Washington, DC: Urban Institute.

Hatch, T., & Grossman, P. (2009). Learning to look beyond the boundaries of representation. *Journal of Teacher Education*, **60**(1), 70–85. doi:10.1177/0022487108328533

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, **23**(1), 27–50. doi:10.1016/j.ecresq.2007.05.002

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**, 1–55. doi:10.1080/10705519909540118

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, **23**(1), 51–68. doi:/10.1016/j.ecresq.2007.09.004

Justice, L. M., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech, Language, and Hearing Research*, **51**(4), 983–1001.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (pp. 17–64). New York: Praeger.

Kerrins, J. A., & Cushing, K. S. (2001). Taking a second look: Expert and novice differences when observing the same classroom teaching segment a second time. *Journal of Personnel Evaluation in Education*, **14**, 5–24. doi:10.1023/A:1008152928725

Kersting, N., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, **61**(1–2), 172–181. doi:10.1177/0022487109347875

Koran, M. L., Snow, R. E., & McDonald, F. J. (1971). Teacher aptitude and observational learning of a teaching skill. *Journal of Education Psychology*, **62**(3), 219–228. doi:10.1037/h0031142

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**(1), 159–174.

Loeb, S., Kalogrides, D., & Béteille, T. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, **7**(3), 269–304. doi:10.1162/EDFP_a_00068

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, **15**(1), 47–70. doi:10.1177/014662169101500106

Mashburn, A. J., Downer, J. T., Hamre, B. K., Justice, L. M., & Pianta, R. C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science*, **14**(4), 179–196. doi:10.1080/10888691.2010.516187

Mashburn, A. J., Downer, J. T., Rivers, S. E., Brackett, M. A., & Martinez, A. (2014). Improving the power of an efficacy study of a social and emotional learning program: Application of generalizability theory to the measurement of classroom-level outcomes. *Prevention Science*, **15**(2), 146–155.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, **79**(3), 732–749. doi:10.1111/j.1467-8624.2008.01154.x

Matheson, A. S., & Shriver, M. D. (2005). Training teachers to give effective commands: Effects on student compliance and academic behaviors. *School Psychology Review*, **34**(2), 202–219.

McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, **28**(5–6), 411–426. doi:10.1016/j.appdev.2007.06.010

McClelland, M. M., Cameron, C. E., Wanless, S., & Murray, A. (2007). Executive function, behavioral self-regulation, and social-emotional competence: Links to school readiness. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on research in social learning in early childhood education* (pp. 83–107). Charlotte, NC: Information Age.

MET Project. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching project*. Seattle: Bill & Melinda Gates Foundation.

Metzger, S. A., & Wu, M. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, **78**(4), 921–940.

Miller, K. F. (2011). Situation awareness in teaching: What educators can learn from video-based research in other fields. In M. G. Sherin, V. R. Jacobs, R. A. Phillipp, (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 51–65). New York: Taylor & Francis.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York: Guilford.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (6th ed.). Los Angeles: Authors.

Neitzel, C., Alexander, J. M., & Johnson, K. E. (2008). Children's early interest-based activities in the home and subsequent information contributions and pursuits in kindergarten. *Journal of Educational Psychology*, **100**(4), 782–797. doi:10.1037/0022-0663.100.4.782

Nicholson-Crotty, S., & Staley, T. (2012). Competitive federalism and race to the top application decisions in the American states. *Educational Policy*, **26**(1), 160–184. doi:10.1177/0895904811428974

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., . . . Vander Borght, M. (2012). Supplementing literacy instruction with a media-rich intervention: Results of a randomized controlled trial. *Early Childhood Research Quarterly*, **27**(1), 115–127. doi:10.1016/j.ecresq.2011.07.002

Pianta R. C., Barnett W. S., Burchinal M. R., & Thornburg K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest*, **10**(2), 49–88. doi:10.1177/1529100610381908

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science*, **315**(5820), 1795–1796. doi:10.1126/science.1139719

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, **38**(2), 109–119. doi:10.3102/0013189X09332374

Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore: Brookes.

Raab, M., Dunst, C. J., Wilson, L. L., & Parkey, C. (2009). Early contingency learning and child and teacher concomitant social-emotional behavior. *International Journal of Early Childhood Special Education*, **1**(1), 1–14.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, **6**(1), 43–74. doi:10.1162/EDFP_a_00022

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, **28**(1), 63–88. doi:10.3102/00028312028001063

Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, **10**(2), 123–140. doi:10.1007/s10857-007-9029-9

Schank, R. (1997). *Virtual learning*. New York: McGraw-Hill.

Seidel, T., Sturmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teachers and Teaching*, **27**(2), 259–267. doi:10.1016/j.tate.2010.08.009

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Taylor & Francis.

Sherin, M., & van Es, E. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, **13**(3), 475–491.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, **15**(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, **57**(1), 1–23.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, **24**(3), 97–117.

Star, J. R., Lynch, K. H., & Perova, N. (2011). Using video to improve mathematics' teachers' abilities to attend to classroom features: A replication study. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (eds.), *Mathematics teachers' noticing: Seeing through teachers' eyes* (pp. 117–133). New York: Routledge.

Stephenson, J., Carter, M., & Arthur-Kelly, M. (2011). Professional learning for teachers without special education qualifications working with students with severe disabilities. *Teacher Education and Special Education: Journal of the Teacher Education Division of the Council for Exceptional Children*, **34**(1), 7–20. doi:10.1177/0888406410384407

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge, MA: Cambridge University Press.

Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, **10**(4), 571–596.

Van Es, E. A., & Sherin, M. G. (2006). How different video club designs support teachers in "learning to notice." *Journal of Computing in Teacher Education*, **22**(4), 125.

van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in Head Start preschoolers with language impairment using scripted book-sharing discussions. *American Journal of Speech-Language Pathology*, **15**(1), 85–95. doi:10.1044/1058-0360(2006/009)

Veenman, M. V., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science*, **33**(3), 193–211. doi:10.1007/s11251-004-2274-8

Walsh, B. A., & Blewitt, P. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal*, **33**(4), 273–278. doi:10.1007/s10643-005-0052-0

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, **103**(2), 455–469. doi:10.1037/a0023067

Wiens, P. D., Hessberg, K., LoCasale-Crouch, J., & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers knowledge related to effective teaching. *Teaching and Teacher Education*, **33**, 24–33. doi:10.1016/j.tate.2013.01.010

Zhang, M., Lundeberg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education*, **27**(2), 454–462. doi:10.1016/j.tate.2010.09.015