Studies in Educational Evaluation 45 (2015) 27–36

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

SCIIENCE: The creation and pilot implementation of an NGSS-based instrument to evaluate early childhood science teaching

Joan N. Kaderavek, Tamala North, Regina Rotshtein, Hoangha Dao, Nicholas Liber, Geoff Milewski, Scott C. Molitor *, Charlene M. Czerniak

University of Toledo, United States

ARTICLE INFO

Article history:
Received 17 September 2014
Received in revised form 9 March 2015
Accepted 10 March 2015
Available online 2 April 2015

Keywords:
Discourse analysis
Teacher assessment
Language of science in classrooms
Validity/reliability

ABSTRACT

This paper describes the development, testing and implementation of the **S***ystematic* **C***haracterization of* **I***nquiry* **I***nstruction in* **E***arly Lear***N***ing* **C***lassroom* **E***nvironments* (SCIIENCE). The SCIIENCE instrument was designed to capture best practices outlined in the National Research Council's *Framework for K-12 Science Education* as they occur within a science lesson. The goals of the SCIIENCE instrument are to (a) assess the quality of science instruction in PK-3 classrooms, (b) capture teacher behaviors and instructional practices that engage students in the lesson, promote scientific studies, encourage higher-level thinking, and (c) provide a feedback mechanism for guiding professional development of PK-3 teachers. Science educators can apply this instrument to teacher behaviors and use the data to improve classroom inquiry instructional methodology.

Introduction

The Next Generation Science Standards (NGSS) and the Framework for K-12 Science Education [Framework] indicates that K-12 classroom instruction should focus on the intersection of scientific and engineering practices, disciplinary core ideas, and crosscutting concepts (National Research Council, 2012). High-quality science instruction should focus on teaching "how we come to know what we know" instead of only teaching "just what we know."

Preschool and early childhood science are overlooked as the necessary foundation for eventually achieving high quality instruction (Pratt, 2007); and yet, science especially lends itself to inquiry, exploration, and curiosity essential for establishing young children's positive attitudes towards school in general as well as towards reading, mathematics, and of course science. There is an unwritten expectation that students will naturally develop an interest in science when it is introduced in middle school or even later in junior high (Keeley, 2009). Furthermore, there is a need for early childhood science if our nation expects to improve science education at subsequent grade levels (McCormack, 2010). Eventual achievement levels in science begin in kindergarten and first grade (Chapin, 2006). Yet, many early childhood teachers are intimidated by science and not well prepared to teach science in early grades (Wenner, 1993).

Thus, to achieve the goal of having US children perform at the highest level in scientific inquiry and knowledge, a number of activities must occur that are focused on early childhood science. First, teacher trainers tasked with developing professional development [PD] in line with the current science Framework must explicitly describe what high quality science instruction looks like in early grades. Without this consensus, and an instrument to measure a teacher's level of achievement in implementing the targeted science instructional practices, researchers will not be able to determine whether a teacher will change his or her science instruction as a result of PD, document whether teaching behaviors will stay consistent over time, or determine what specific instructional practices contribute to an increase in child knowledge or skill level.

This paper describes the development, testing and application of an instrument known as the **S***ystematic* **C***haracterization of* **I***nquiry* **I***nstruction in* **E***arly Lear***N***ing* **C***lassroom* **E***nvironments* (SCIIENCE). The SCIIENCE instrument was designed to objectively capture the presence and frequency of specific best practices outlined in the Framework as they occur within a science lesson and focuses exclusively on teacher behaviors. The goals of the SCIIENCE system are to (a) provide a standardized instrument for

* Corresponding author at: University of Toledo, 2801 W. Bancroft Street, MS 311, Toledo, OH 43606-3390, United States. Tel.: +1 419 530 8040.
E-mail address: scott.molitor@utoledo.edu (S.C. Molitor).

http://dx.doi.org/10.1016/j.stueduc.2015.03.003
0191-491X/

assessing the quality of science instruction in a classroom setting for children grades PK-3, (b) capture the instructional practices that engage students in the lesson, promote scientific studies, encourage higher-level thinking, and (c) provide feedback for guiding professional development of PK-3 teachers.

Beyond this focus, our intent is to create an instrument that is standardized, comprehensive, geared toward early childhood classrooms and psychometrically sound. A standardized instrument is defined as an assessment that offers consistent procedures and uniform application and has the potential to compile and compare findings across teachers and different science lessons. A comprehensive instrument is needed because researchers and educators need to capture a wide range of adult behaviors as they occur within science instruction. Furthermore, the most frequently used observational instruments – Inside the Classroom Observation and Analytic Protocol (ITC COP; Horizon Research, Inc., 2000) and the *Reform Teaching Observation Protocol* (RTOP; Sawada et al., 2000) have limitations.

The Horizon rating system provides a global view of various aspects of science instruction but does not permit a fine-grained analysis of specific teacher practice (Henry, Murray, Hogrebe, & Daab, 2009; Henry, Murray, & Phillips, 2007). The RTOP also has limitation in that it uses a Likert scale to assess the quality of classroom instruction. Teachers can be challenged to understand what specifically they need to do to improve their overall instructional quality rating (i.e., to move from a level 4 to a level 5) when a Likert scale is used to evaluate the quality of teaching. The SCIIENCE instrument was developed to surmount these limitations.

The following sections in this paper elaborate the development, testing, and application of the SCIIENCE instrument at the project-level. The authors will describe: (a) the theory and prior research that informed the development of an instrument for evaluating teachers' instructional practice; (b) the theory and background that informed the development of the SCIIENCE instrument; (c) the individual codes and the way in which the SCIIENCE instrument is used to evaluate the quality of science inquiry instruction; (d) project-level data documenting the reliability and validity of the SCIIENCE instrument and the comprehensive plan developed to further document the instrument's validity; (e) an application of this instrument to improve teacher professional development; and (f) a discussion of the limitations of the instrument and future research to be completed using the SCIIENCE instrument.

Theory and background of instruments for evaluating classroom quality and instructional methods

There is a scarcity of literature that examines early childhood science education. But we can look at how teachers approach science inquiry with the hope of fostering good science inquiry technique and the benefit of making teachers more comfortable with teaching science.

Teachers play a vital role in encouraging young children to engage in sophisticated behaviors and verbal interactions (De Kruif, McWilliam, Ridley, & Wakely, 2000). The Measures of Effective Teaching project found that teachers identified as more effective caused students to learn more and teachers identified as less effective caused students to learn less (Cantrell & Kane, 2013). We believe that investing in practices and policies that support effective teaching will result in improved science inquiry. However, we also believe that teaching and learning is too complex for a single measure of performance. To identify best practices, multiple measures are required.

Current instruments that assess quality in early childhood settings include the Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980), now the Early Childhood Rating Scale – Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). This instrument, which consists of 43 items arranged in 7 subscales, has been widely used in the early education field to assess the quality of the preschool and kindergarten learning environments in childcare centers. The ECERS-R scale consists of 43 items arranged in seven subscales. The subscales are assigned a score from 1 to 7 and an overall scale score is calculated by averaging the subscale scores. A score of 1 indicates inadequate quality, 3 indicates minimal quality, 5 indicates good quality and 7 indicates excellent quality. The ECERS-R assesses the quality of classroom routines, the quality of the activities, the availability of materials, provisions for parents and staff, and the interaction between teachers and children. In addition to assessment, the scale is used in lesson planning and professional development (Tout et al., 2010). Though the ECERS-R has a proven track record of success and is commonly used in informal child care settings, it was not designed to be used in classroom environments with a more academic focus. The ECERS-R only dedicates one item out of 43 to science.

A similar scale to the ECERS-R is the School-Age Care Environment Rating Scale (SACERS; Harms, Jacobs, & White, 1996). The SACERS scale consists of 49 items arranged in seven subscales. Like the ECERS-R, the subscales are assigned a score from 1 to 7 and an overall scale score is calculated by averaging the subscale scores. A score of 1 indicates inadequate quality, 3 indicates minimal quality, 5 indicates good quality and 7 indicates excellent quality. Though the SACERS includes environments serving children from age 5 to 12, like the ECERS-R, it was designed to assess informal child care environments and only dedicates one item to science.

Like the ECERS-R and SACERS, the Classroom Assessment Scoring System (CLASS; Pianta, Karen, Paro, & Hambre, 2008) is a commonly used instrument to observe teacher practices in early childhood classrooms. The CLASS is comprised of three domains and ten dimensions. The dimensions are scored on a scale from 1 to 7. A score of one signifies that all, or almost all, indicators in the low range are present. A score of 7 signifies that all, or almost all, indicators in the high range are present. The dimensions scores are averaged accordingly for an overall domain score in each of the three domains. Unlike the ECERS-R and the SACERS, the CLASS is appropriate for assessing informal childcare settings as well as formal classrooms by measuring teacher performance in infant settings through secondary grades. The focus of the CLASS is to rate the teacher–student interaction in the domains of emotional support, classroom organization, and instructional support. Though these domains play an important role in quality science inquiry, the CLASS does not examine a science-specific dimension.

The Early Language and Literacy Classroom Observation (ELLCO; Smith & Dickinson, 2002) is an assessment instrument used in preschool through third grade classrooms. The ELLCO is comprised of three parts: a literacy environment checklist, a general classroom observation and teacher interview and a literacy activities rating scale. The ELLCO has a proven track record of increasing literacy behaviors in children through the modification of the environment and teacher mediation (Wayne, DiCarlo, Burts, & Benedict, 2007).

Assessing the learning environment is imperative, but equally important is the assessment of the instruction provided to students (Koehler-Hak, 2008). We propose that a science instrument has potential for improving science instructional practices because other instruments such as CLASS, ELLCO, ECERS-R and SACERS have been used effectively to highlight issues in instructional domains such as language and literacy. For example, the CLASS has been used to identify how teacher instructional practices impact children's academic outcomes; specifically noting the positive impact of teachers' use of high-level inferential language (Howes et al., 2008).