7-2015

A Written Instrument for Assessing Students' Units Coordination Structures

Anderson Norton
Virginia Tech

Steven Boyce
Portland State University

Nathan Phillips
Virginia Tech

Tessa Anwyll
Virginia Tech

Catherine Ulrich
Virginia Tech

See next page for additional authors

Authors

Anderson Norton, Steven Boyce, Nathan Phillips, Tessa Anwyll, Catherine Ulrich, and Jesse L. M. Wilkins

A Written Instrument for Assessing Students' Units Coordination Structures

Anderson Norton
Virginia Tech, USA
Steven Boyce
Portland State University, USA
Nathan Phillips, Tessa Anwyll, Catherine Ulrich & Jesse L. M. Wilkins
Virginia Tech, USA

Units coordination refers to students' abilities to create units and maintain their relationships with other units that they contain or constitute. In recent research, units coordination has arisen as a key construct that mediates opportunities for student learning across several domains of mathematics, including fractions knowledge and algebraic reasoning. To date, assessments of students' stages of units coordinating ability have relied upon clinical interviews or teaching experiments whose time-intensive nature precludes opportunities for conducting large-scale studies. We introduce a written instrument that teachers and researchers can use with large populations of students. We report on the reliability and validity of assessments based on the instrument.

Keywords: assessment; fractions; multiplicative reasoning; units coordination; written instrument

INTRODUCTION

Throughout their schooling experiences, students' abilities to build and coordinate units mediate ways in which they can engage in mathematics. In the earliest years, children's number sequences determine ways they can count. Steffe (1992) defined a number sequence as "a sequence of abstract unit items that contain records of counting" (p. 263), and he identified the following progression in children's number sequences. With an *Initial Number Sequence* (INS), 1 is a unit that can be counted, but the record of counting (e.g., "one, two, three, four") is not itself a unit. Rather, it symbolizes the activity of counting 1s from which the child might *count on* ("five, six, seven"). Once students unitize their records of counting, they can begin coordinating two levels of units—the 1s being counted and the composite unit containing those 1s. With a *Tacitly Nested Number Sequence* (TNS), students can make this coordination in the activity of counting, enabling them to *double count* (e.g., three 4s as "one, two, three, four; that's 1. Five, six, seven, eight; that's 2. Nine, ten, eleven, twelve; that's 3"). Within the *Explicitly Nested Number Sequence* (ENS),

Correspondence: Anderson Norton,
Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, United States of America.
E-mail: norton3@vt.edu

students can iterate the composite unit to produce a unit of units of units *in activity* (e.g., building up 12 as three units of four 1s by physically or imaginatively iterating a composite unit of four 1s). Once a student can *assimilate* such three-level unit structures, without the need to build them through activity, the student has constructed the operations that undergird a *Generalized Number Sequence* (GNS).

In later years, the levels of units students coordinate afford and constrain their abilities to reason multiplicatively (Steffe, 1994; Hackenberg, 2010; Tillema, 2013), conceptualize fractions (Hackenberg & Tillema, 2009; Hunting, 1983; Norton & Boyce, 2013; Steffe & Olive, 2010), add integers (Ulrich, 2012), determine geometric area (Clements, Battista, Sarama, & Swaminathan, 1997; Reynolds & Wheatley, 1996), and reason algebraically (Ellis, 2007; Olive & Çağlayan, 2007). Steffe (2002) articulated the link to fractions in his reorganization hypothesis, which stipulates that students can reorganize their ways of operating with whole number units in order to work with fractional units and construct knowledge of fractions. For example, just as a student with an ENS can partition a unit of 12 into three 4s, disembed a unit of 4 from 12, and iterate the unit of 4 to produce 8, she can learn to partition a whole into three thirds, disembed one of those thirds from the whole, and iterate it to produce 2/3. Such students might even produce an improper fraction, say 5/3, through iteration of a unit fraction, but conceptualizing 5/3 as a unit of five units of 1/3 would require the student to assimilate three levels of fractional units (1/3, the whole, and 5/3), just as the GNS involves assimilating three levels of whole number units (Hackenberg, 2007).

Because students' units coordinating ability has such strong implications for their learning of whole numbers, fractions, integers, and algebra, it is important that teachers and researchers have instruments for assessing this ability. To date, researchers have relied on teaching experiments and clinical interviews to assess students' abilities to coordinate various levels of units—a time-intensive process that precludes large-scale study. The purpose of this paper is to present a written instrument for assessing the psychological structures students have available for coordinating units. The instrument utilizes tasks similar to those used in teaching experiments and clinical interviews, along with a rubric developed through the present study. We report on instrument design, as well as initial measures for validity and reliability of assessments using the instrument. Note that we do not report on a large-scale study here, but rather a validation study of a written instrument that would allow for large-scale studies. With the reliability and validity of the instrument established, other researchers can use the instrument in large-scale studies related to units coordination.

We begin with a review of literature related to assessing units coordination and the theoretical basis for our task design. Then we describe the methods we used to design the rubric and to test the instrument as a tool for assessing students' units coordination structures. We present the instrument along with results regarding validity and reliability. Finally, we discuss the implications of our design in terms of the instrument's utility for teaching and research.

THEORETICAL FRAMEWORK

Units coordination refers to the ways students can build and work with various levels of units (Steffe, 1992). We can characterize a progression in students' abilities to coordinate units in terms of stages, as determined by the number of levels of units they can assimilate (cf., Hackenberg, 2010). Consider Figure 1 and potential responses to the following related task: "If the small red bar fit into the medium yellow bar three times, and the medium yellow bar fit into the long blue bar four times, how many times would the small red bar fit into the long blue bar?"

Figure 1. Bars task

Table 1. Stages of units coordination

	Students' Unit Structures	**Students' Reasoning on the Bars Task**
Stage 1	Students can take one level of units as given, and may coordinate two levels of units in activity.	Students mentally iterate the short bar, imagining how many times it would fit into the longer bar. This activity might be indicated by head nods or sub-vocal counting.
Stage 2	Students can take two levels of units as given, and may coordinate three levels of units in activity.	Students mentally iterate the medium bar four times, with each iteration representing a 3. This activity might be indicated by the student uttering "3, 3, 3, and 3; 12."
Stage 3	Students can take three levels of units as given, and can thus flexibly switch between three-level structures.	Students immediately understand that there are four threes in the long bar. This assimilation of the task might be indicated by an immediate response of "12," buttressed by an argument that 12 is four 3s.

At Stage 1, students can assimilate units of 1 and may even assimilate composite units as a collection of 1s, but the relationship between these two levels of units is not multiplicative; in other words, they do not assimilate the composite unit as a unit that is n times as big as 1. In the bars task, these students are unlikely to use the given relations. Instead, they might take the small red bar as a unit of 1 and iterate its length across the length of the long blue bar to determine the unknown relation figuratively. At Stage 2, students can take two-level multiplicative relations as given. Thus, they can meaningfully assimilate the given relations; however, they have to build up the third level of units through activity. For example, they might treat the medium yellow bar as a unit of three 1s (small blue bars) and mentally iterate that composite unit four times, as specified by the given relation between the medium yellow bar and the long blue bar. At Stage 3, students can assimilate the given relations into a unit of units of units: the long bar as four medium bars, each of which is three small bars. As such, these students can immediately apprehend the unknown relationship as a multiplicative coordination of the given two relationships without the need to perform iterating actions, physically or mentally. Table 1 summarizes the three stages of coordinating units.

The bars task and stage descriptions suggest a method for assessing students' ways of coordinating units. Indeed, we adopted a similar approach in designing our tasks, but in order to establish the face validity of our tasks, we need to review literature on how other researchers have characterized and assessed units coordination.

Assessing units coordination

Steffe's (1992) number sequences correspond to the three stages in Table 1 as follows: Students with an INS and many with a TNS are operating at Stage 1; students with an ENS are operating at Stage 2, and students with a GNS are operating at Stage 3. Within Stage 1, the distinguishing characteristic of TNS students is that they can work with composite units, indicating progress toward Stage 2. With these correspondences established, we can consider the work of Steffe and colleagues, in assessing students' number sequences, as assessments of their stages of units coordination.

Whole number contexts

Steffe's assessments of number sequences occur in the context of teaching experiments (Steffe & Thompson, 2000; Steffe & Ulrich, 2014) where he designs tasks to test hypotheses about how students operate and to provoke new ways of operating. For example, consider the following interaction with a third grade student named "Maya" (Steffe, 1992). Steffe presented Maya with a red bar and several blue bars, produced such that the blue bar would fit into the red bar six times, much like the bars task illustrated in Figure 1. After Maya determined that six blue bars fit into the red bar (by aligning blue bars with the red bar), Steffe placed two orange bars on a single blue bar (fitting exactly) and removed the extra blue bars. Steffe asked Maya how many orange bars would fit into the red bar, instructing her to "figure it out with the blue ones" (p. 279). Maya appeared to respond by sub-vocally uttering the following sequence: "1, 2; 3, 4; 5, 6; 7, 8; 9, 10; 11, 12."

Based on Maya's response, Steffe inferred that Maya could "distribute her unit of 2 across the six units of 1 prior to counting" (p. 280). However, the result was still a single number sequence, based on units of 1. As such, Steffe assessed Maya as operating with a TNS, which he affirmed after posing the following task. Maya had used blocks to make six rows of 3, which Steffe hid with a cover. Steffe then asked Maya to determine how many more rows of 3 she would need to have a total of ten rows of 3. Initially, Maya responded by asking, "You mean you add three blocks to them under there?" (p. 281). When Steffe repeated the question, emphasizing that she was to add *rows* of 3, Maya responded as follows: "One more row and one more. Four more rows—you have six, and you count up to six and then take another three, and that's nine, and then one more and that's ten!" (p. 282). This response indicates that Maya's goal had been to make 10 from 6 without distinguishing levels of units: A row simply meant 3; a block simply meant 1. Although Maya could construct composite units through activity (e.g., six 2s), she did not reliably assimilate units at two levels (e.g., blocks and rows).

Contrast Maya's response to the bars task with that of Johanna—a student operating with an ENS. After about 15 seconds in thought, Johanna responded "12" and explained, "Well, six plus six is twelve, and each two bars fits on one big bar, and that makes 12!" (p. 292). From this and subsequent activity in the teaching experiment, Steffe determined that Johanna could iterate composite units, understanding a composite unit as, say, a unit of 3 and a unit of three 1s, at the same time. This enabled Johanna to produce three levels of units in activity. "By operating, Johanna could establish a unit containing seven units each of which contained three units as the meaning of 'seven rows with three blocks in each row.'" (p. 296).

"In general, I think of units coordinating as the mental operation of distributing a composite unit across the elements of another composite unit" (Steffe, 1992, p. 279). In the case of TNS, this distribution occurs within a single level number sequence; whereas, with the ENS, students produce a correspondence between units across two levels of their number sequence (e.g., "5 is 1; 10 is 2; …"). With the GNS, students can immediately assimilate the results of such distributions as a third level and work flexibly across all three levels—what Steffe refers to as a "recursive units coordinating scheme" (p. 305). The reorganization hypothesis posits a direct connection between these ways of operating with units across whole numbers and fractions: "Children's fractions schemes can occur as accommodations in their numerical counting schemes" (p. 267).

Fractions Contexts

In line with the reorganization hypothesis, researchers have found strong connections between students' ways of operating with whole number units and their ways of operating with fractional units (Boyce & Norton, in review;

Hackenberg, 2007; Hackenberg, 2010; Hackenberg & Tillema, 2009; Olive & Steffe, 2002; Steffe, 2002). As such, students' ways of operating in fractional contexts provide further indication of their stages of units coordination. For instance, in partitioning a whole into *n* equal parts, students are producing a composite unit—the whole as a unit of *n* *n*ths. Although students operating at Stage 1 might carry out this activity, only students operating at Stages 2 and 3 can disembed one of those parts without destroying the whole because they have to simultaneously treat that part as a unit of 1 and 1 of *n* parts in the composite whole (Steffe, 2002).

Relying on the cross-contextual connection, Hackenberg (2007) assessed students' abilities to coordinate units using a fractions task. She asked students to mark off a fair share of a candy bar (represented with a rectangular piece of cardboard) split among six friends. Some of the students were able to mark off a share without partitioning the entire candy bar, and furthermore, were able to justify the appropriateness of the size of the share by iterating it six times within the whole. These actions indicate that the students were operating at Stage 2 or 3 because they could treat the disembedded unit as a unit of 1 without losing its 1-to-6 relation with the whole.

Multiplicative concepts

Building from Steffe's (1994) multiplication schemes, Hackenberg (2010) characterized students' units coordinations in terms of multiplicative concepts. She distinguished three multiplicative concepts, corresponding to the three stages in Table 1: multiplicative concept 1 (MC1) in which students can coordinate two levels of units in activity (Stage 1); multiplicative concept 2 (MC2) in which students can take two levels of units as given and coordinate three levels of units in activity (Stage 2); and multiplicative concept 3 (MC3) in which students can take three levels of units as given (Stage 3). To understand the relation to multiplicative reasoning, consider how students might assimilate the following situation: "At a bakery, a muffin tin has four rows with six muffins in each row" (p. 391). MC1 students form the goal of determining the total number of muffins but have to build the four units of six in the activity of counting (e.g., "6; 7, 8, 9, 10, 11, 12; 13, …, 18; 19, …, 24"). MC2 students can anticipate the four rows of six as a totality—four iterations of a unit of 6—though they may have to carry out units coordination activity to determine the result (e.g., "6 and 6 is 12, and another 12 is 24"). MC3 students retain this three-level structure (24 as four units of 6) in further operating (e.g., "24 is four 6s, so 48 is twelve 4s").

MC3 students and many MC2 students can reverse their multiplicative reasoning in both whole number and fractions contexts (Hackenberg, 2010). A "reversible multiplicative reasoning" (RMR) task in a whole-number context might look like the following: "Sara's stack of CDs is 65 cm tall. That's 5 times the height of Roberto's stack of CDs. Can you draw a picture of this situation? How tall is Roberto's stack?" (p. 402). What makes the task reversible is that the task is iterative in nature ("5 times") and yet students have to partition (or divide) the result in order to find the unknown quantity that is iterated.

An RMR task in a fractions context might look like this: "A 2-foot bar is three times the length of your candy bar. Make your candy bar and tell me how long it is (no erasing the foot-mark)" (p. 410). Carlos—a sixth-grade MC2 student from Hackenberg's (2010) teaching experiment—eventually solved the task by splitting the two feet within the two-foot bar "into little pieces so I can add them all up into three pieces" (p. 411). This response indicates that Carlos knew he needed to reverse the iteration by partitioning the given 2-foot bar, but that he did not know what units to use within each foot—a limitation of MC2 reasoning.

In the present study, we used ideas from Steffe (1992, 2002) and Hackenberg's (2007, 2010) teaching experiments to design our tasks, and our assessments of units

coordination align with their assessments of number sequences and multiplicative concepts. However, we frame students' ways of operating with units in terms of units coordination *structures*. For that reason, we turn to Piaget's structuralism.

Units coordinating structures

In Piaget's (1970a) epistemology, mathematical operations are mental actions that have been organized within a system for composing and reversing them, where "reversibility is defined as the permanent possibility of returning to the starting point of the operation in question" (Inhelder & Piaget, 1958, p. 272). Mathematical structures are the systems of transformation that organize operations (Piaget, 1970b). For example, in early childhood, we develop a structure that organizes translations of objects in space. Translations arise as mental actions when we internalize physical actions of moving objects or tracing the movements of objects with our eyes. Translations become mathematical operations once they are organized in such a way that (1) they can be composed to form other translations and (2) each translation can be negated by a corresponding translation (reversibility). The result is a structure for displacements, which can be modeled by an algebraic group, isomorphic to vector addition in three-dimensional Euclidean space (Piaget & Inhelder, 1967/1948).

Like algebraic groups, structures comprise the following three ideas: "the idea of wholeness, the idea of transformation, and the idea of self-regulation" (Piaget, 1970b, p. 5). *Wholeness* refers to the idea that, although structures have individual elements, the relationships between elements are defined by rules that apply to a consolidated totality, rather than particular elements. *Transformation* refers to the idea that structures are not static forms; structures frame "intelligible change that always preserves invariance in certain respects" (p. 20). *Self-regulation* refers to the idea that the "transformations inherent in a structure never lead beyond the system, but always engender elements that belong to it and preserve its laws" (p. 14). Collectively, these three ideas correspond to the closure of a group under composition of its elements. In the case of mathematical structures, there is also reversibility, corresponding to the idea that every element of a group has an inverse element so that composition of the two elements yields the identity (p. 15).

The elements of mathematical structures are operations (e.g., the mental actions of translating objects in space, organized in the displacement group). Whereas Steffe (1992) characterized units coordinating as distributing one composite unit across another, we can describe the operations of a units coordinating structure in greater detail. They include unitizing, iterating, partitioning, and disembedding (Steffe & Olive, 2010). Figure 2 represents the form of a units coordinating structure, which might be produced by iterating a composite unit. Units coordinating structures are ways of operating within such forms, especially regarding levels of units. For example, with regard to Figure 2, a student operating at Stage 3 might disembed one of the smallest units from the unit of 24, iterate that unit 3 times, unitize the result as a unit of three 1s, and then iterate that composite unit eight times to reproduce 24 as a unit of eight units of three 1s (three levels of units). The structural aspect of units coordination refers to the idea that such organizations of operations completely define and regulate possible and necessary relationships between units.
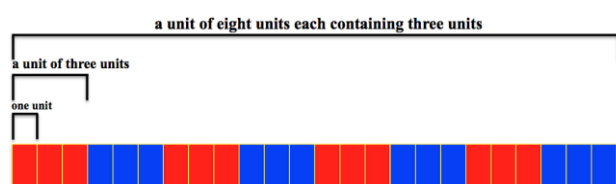

Figure 2. Representation of a structure for coordinating units

Here, we describe the units coordinating structures available to students at each stage.

Although a student operating at Stage 1 might produce composite units, as Maya did (Steffe, 1992), the resulting structure would not be 24 as a unit of eight units of three 1s but rather a chain of 1s segmented into 3s, and she would not be able to work with units of 3 and units of 1 simultaneously. She could iterate 1 three times or 24 times, and she could partition the chain of 1s into sets of 3. She could even carry out the activity of iterating a partitioned set of 3 eight times, but she could not unitize three 1s as one 3 and iterate it to simultaneously produce 24 as 24 1s and eight 3s. In sum, the Stage 1 student can carry out iterating and partitioning activity, always working from a unit of 1. The ideas of wholeness, transformation, and self-regulation inherent in the structure are captured by the inverse relationship between the activities of iterating a unit of 1 and partitioning a collection into units of 1.

Students operating at Stage 2 can unitize three 1s as one 3 and establish a multiplicative 3-to-1 relationship between those two units. As a result, these students can produce the form illustrated in Figure 2 by iterating the composite unit of 3. Students at Stage 2 can also disembed a unit of 1 from a unit of 3, maintaining its constituency as part of the 3, even while treating it as its own unit. In this sense, unitizing and disembedding behave like inverse operations (similar to partitioning and iterating). These students can also disembed a 3 from the eight 3s that comprise 24. However, when iterating these disembedded units, students do not maintain the new units they produce. For example, a student might disembed a unit of 3 and iterate it two times to produce 6 as two 3s, but when asked how many 3s are in four 6s, the student will not be able to take for granted 6 as a unit of two 3s (e.g., "four 6s is four sets of two 3s, so the answer is 8"); he will have to build up from 3 again (e.g., "one, two 3s is 6, then two more, and two more, and two more, so eight 3s is four 6s").

Students at Stage 3 do maintain the third level of units. In fact, they can assimilate situations involving three levels of units into their units coordinating structures, accounting for and distinguishing all three levels at once, without having to carry out any activity. For example, consider the following task: "A teacher has a classroom with six rows of four desks but needs to add 12 more desks to the classroom. How many total rows will there be if he decides to keep the desks in rows of four?"[1] Students at Stage 3 can immediately assimilate the six rows of 4 as a unit of units of units. Even if they do not know the total number of desks, they know there are six units at the second level and that each of these units contains four units at the first level (units of 1). Thus, they can focus on the problem of determining how many units of 4 are created from the 12 extra desks. In contrast, students at Stage 2 generally become consumed with the (unnecessary) task of computing 6 times 4 and, in the process, lose track of the stated goal (Boyce & Norton, in review).

Units coordination structures can take different forms for different contexts, but the constituent operations are the same. Indeed, this is the basis for Steffe's (2002) reorganization hypothesis. In particular, consider Figure 3, which represents the form of a structure for coordinating three levels of units within an improper fraction. Here, 8/3 is a unit of eight units of 1/3, three of which make the whole unit. Similar to the way they operate with whole number units, students operating at Stage 3 can partition the whole into three parts, disembed one of those parts and iterate it eight times to produce 8/3 while maintaining the 3-to-1 relationship between the iterated part (1/3) and the whole (Hackenberg & Lee, 2015).

---

[1] Thanks to Amy Hackenberg for developing the original version of this task.
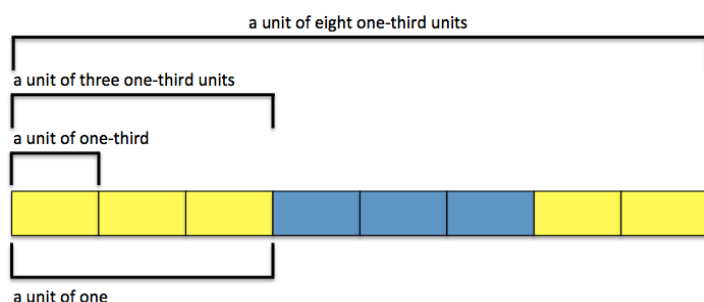
Figure 3. 8/3 as a unit of eight units of 1/3, 3 of which make the whole

RESEARCH DESIGN

Instrument design and testing occurred in three phases. In Phase I, the first and second authors designed tasks based on our theoretical framework and previous studies on students' units coordinations. Then, this pair of raters (R1 and R2) conducted clinical interviews with 20 middle school students to design a rubric for assessing student responses to those tasks. In Phase II, a second pair of raters (R3 and R4; third and fourth authors) used the instrument—tasks and rubric—to score the level of units coordination for the 20 students. The validity and reliability of these scores were then assessed. In Phase III, the first and second pairs of raters worked to improve the rubric, and then we recruited a third and fourth pair of raters (R5-R8) to score the level of units coordination for the 20 students. The validity and reliability of the scores based on the revised rubric was then evaluated.

Tasks

In previous studies, we have used bars tasks to assess students' units coordinating activity during clinical interviews (Boyce & Norton, in review; Norton, Boyce, Ulrich, & Phillips, 2015). We designed these tasks such that students could readily assimilate them into available units coordination structures. The interviewer would give the student three strips of construction paper ("bars") to manipulate, as illustrated in Figure 2. Then, the interviewer would ask the student to determine how many times the small bar fits into the medium bar and how many times the medium bar fits into the long bar; students would generally respond by sliding the small (medium) bar along the length of the medium (long) bar. Finally, the interviewer would ask the student to use those relationships to determine how many times the small bar would have to fit into the long bar.

Subsequent bars tasks used a different set of bars, and rather than asking the student to determine the first two relationships, the interviewer would ask the student to pretend that they have particular relationships, disproportionate to the actual lengths of the bars. Then, the interviewer would ask the student to determine the third relationship based on the given ones. The interviewer would use the new set of bars to ask this kind of question in three different ways, using all three combinations of relationships: (1) where the relationship between the small bar and the long bar is unknown; (2) where the relationship between the small bar and the medium bar is unknown; and (3) and where the relationship between the medium bar and the long bar is unknown.

We designed tasks for the written instrument based on the bars tasks we had used during clinical interviews. In fact, Tasks 1-6 in the written instrument (see Appendix A) also provide illustrations for the clinical interview tasks described above. We included Task 7 to assess how a student might operate when the unknown relationship cannot be represented by a whole number. Collectively, these

seven tasks are intended to elicit written responses from students that would provide indication of their available structures for coordinating units.

Data collection

To design the rubric, we conducted clinical interviews with 20 sixth-grade students. Steffe (2007) has estimated that, by the end of fifth grade, 30-50% of students in the United States operate at Stage 1, with the remainder of the population operating at Stages 2 or 3. Thus, sixth-grade students constitute an ideal population for our study because we might expect roughly one third of them at each stage. The 20 students in our study came from a high-needs rural school in the Southern United States; 63% of the students at the school receive free or reduced price lunch. The students were enrolled in one of five sections of a sixth-grade mathematics class taught by the same teacher.

Clinical interviews occurred over a three-day period in August of 2013. R1 and R2 independently conducted interviews with individual students using the written instrument tasks (Appendix A). Each interview lasted about 15 minutes. We handed students tasks on a sheet of paper, one at a time (with the exception of the first three tasks, which are clustered together), and asked the students to respond in writing. After the students had recorded their responses on the paper, we asked the students to explain what they had written and why they had written it, without them writing anything more on their papers. Then, we moved on to the next task, on a new sheet of paper, and proceeded in the same manner through all seven Tasks (five sheets of paper; Appendix A). We video-recorded the clinical interviews so that we could use them for retrospective assessment of students' units coordination structures. However, R1 and R2 conducted their first assessments relying on written responses alone. Each rater independently rated all written responses.

Data analysis

R1 and R2 used the 20 clinical interviews, along with students' written responses, to design and test the instrument. First, we assessed students' units coordination structures in two ways: based on written responses alone, and based on video-recorded actions (including verbalizations) in conjunction with written responses. We then used these assessments to design a rubric that other raters could use. Finally, we measured criterion-related validity and inter-rater reliability for assessments conducted by additional raters using this rubric and a revised version of it.

Phase I: Rubric design

R1 and R2 had no rubric for conducting their assessments of students' written responses. Rather, each rater used his knowledge of units coordination and considered student responses holistically in order to identify the stage of operating that best explained the students' responses. With their written assessments complete, these two raters independently reviewed video-recordings for the 20 clinical interviews and conducted a second assessment based on student actions in the video, in conjunction with the corresponding written responses.

While conducting their assessments based on video-recorded responses in conjunction with written responses, the raters looked for student actions that might provide further indication of how students had operated and how this was indicated in their written responses. They paid particular attention to cases where making a final inference had been particularly challenging based on written responses alone. R1 and R2 gathered indicators for each stage of units coordination in order to build the rubric described in the next section. These indicators could be present in students' written responses but were correlated with actions (including

verbalizations) observed in the video that provided stronger indication of each stage of operating.

Phase II: Measuring validity and reliability

With their video-based assessments complete, R1 and R2 reconciled their scores to achieve a final assessment of students' stages of units coordination, which might differ from assessments based on written responses alone. This final assessment was taken as the "true score" for all subsequent rater comparisons. To test reliability and validity of the instrument, we recruited a second pair of raters (R3 and R4). The new raters independently assessed students' stages of units coordination based on written responses and using the rubric. They used the rubric (similar to the one in Appendix B) to score the 20 written assessments following the instructions below.

1. For each student, set of tasks (Tasks 1-3, Tasks 4-5, Task 6, and Task 7), and stage, each rater noted which indicators seemed present in the student's response; as such, each indicator was treated as a dichotomous item, scored as either 0 or 1.

2. For each student and set of tasks, each rater determined which stage (or Stages in the case of Tasks 1-3) seemed to fit best with the indicators noted.

3. After completing these within-task assessments, each rater looked across all seven tasks (four sets of tasks) to determine an overall assessment for each student (Stage 1, 2, or 3).

4. Each rater also noted any difficulty (s)he had in using the rubric for a particular student/task/stage; difficulties might occur at the indicator level, the task level, or in the overall assessment.

When R3 and R4 finished scoring, we computed Cohen's kappa, as a measure of inter-rater reliability (Cohen, 1960). We also computed Kendall's tau and Spearman's rho between each of the two new raters' scores and the final assessment from the first two raters' reconciliation. This correlation provided a measure of criterion-related validity for the scores from the instrument.

In order to test the reliability of the instrument with a different population, the first pair of raters (R1 and R2) used the rubric to score a second set of data. This data came from another project that used the same items. Students in the population were enrolled in seventh grade at another high-needs school about 100 miles from the first school. We randomly selected 20 students from that population and scored their responses based on the rubric. Once again, we used Cohen's kappa (weighted) as a measure of inter-rater reliability.

Phase III: Measuring validity and reliability of the revised instrument

As reported in the results section, Kendall's tau and Spearman's rho for R3 and R4 were high as were kappa scores for R1 and R2 with the second set of data. However, rater agreement, based on kappa scores for R3 and R4, was only moderate. Based on our conversations, we felt that this low reliability was due to ambiguities in the second pair of raters' readings of the rubric. All four raters had kept notes on their use of the rubric with their respective sets of data. Notes included which indicators were used and any difficulties raters had in identifying indicators or stages of units coordination. We used these notes to revise the indicators in the rubric. Then we recruited a third and fourth pair of raters (R5-R8) to use the revised rubric with the original set of data. We purposefully selected one pair of researchers who had familiarity with the construct of units coordination (R5 and R6) and one pair of prospective teachers who did not (R7 and R8). We report results of both iterations in the next section.

RESULTS

Results concern the design of the instrument itself, as well as measures of validity and reliability. We begin by considering face validity of the tasks alone. Then we share results concerning rubric design. Finally, we report results from measuring the validity and reliability of the scores associated with the instrument as a whole (including the rubric).

Face validity of the tasks

Using bars tasks like the one illustrated in Figure 2, the first two authors have engaged in multiple teaching experiments and clinical interviews to assess students' ways of coordinating units (Boyce & Norton, in review; Norton, Boyce, Ulrich, & Phillips, 2015). Items for the written instrument are based on such tasks and align with the theoretical framework we have presented. In particular, the first three tasks (see Appendix A) align with the teaching experiment tasks that Steffe (1992) used to assess students' number sequences, which correspond to stages of units coordination. The fourth task demands similar ways of operating, except students cannot rely on the drawn figures to determine the unknown relationship because the bars are not drawn to scale. As such, the task might help raters determine whether students are operating at Stage 1 and not using the given relations (see Table 1).

Task 4 also aligns with Hackenberg's (2007, 2010) construct of students' multiplicative concepts: "the units coordinations that they can take as given prior to activity" (2007, p. 27). Students at Stages 2 and 3 should be able to take the two stated relationships as given and use them to determine the third relationship multiplicatively, though Stage 2 students would need to do this through activity (e.g., mentally iterating and counting six 2s). Furthermore, Tasks 5 and 6 require students to reverse their multiplicative reasoning to determine the unknown relationship (Hackenberg, 2010). Finally, Task 7 incorporates reversible multiplicative reasoning in a setting where the unknown relationship is not a whole number. Based on Steffe's (2002) reorganization hypothesis, students' ways of operating with whole number units should be available for operating with fractional units. Student responses to Task 7 could distinguish students operating at Stages 2 and 3 by indicating how they account for the extra unit—the leftover unit fraction.

Descriptive statistics also support face validity of the tasks (see Table 2). Without a rubric, R1 and R2 were able to use students' written responses to the tasks to assess the students' units coordination structures. They achieved 80% agreement in their scores (16 out of 20) indicating that, based on their knowledge of units coordination, the raters were able to use student responses to the tasks to assess units coordination structures, even without a rubric. Moreover, there is general agreement across these assessments, the raters' assessments using the video-recorded clinical interviews, and the reconciled assessments using the clinical interviews. As described in the next section, we used discrepancies between these scores to inform the design of a rubric that might support greater validity and reliability of rater scores.

Rubric

In 13 of the 20 cases presented in Table 2, we find perfect agreement across all four independent scores. The first two authors used these cases to explicitly identify indicators for the rubric. For example, some students' written responses indicated that they were not using the given relations to determine the unknown relation but were simply relying on the appearance of the given bars; this was a strong indicator that these students were operating at Stage 1. We captured such indicators in the

Table 2. First pair of raters' assessments of units coordination

Student#	Written R1	Written R2	Video R1	Video R2	Video Final
1	3	2	2	2	2
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	2	2	2	2	2
7	1	1	1	1	1
8	2	2	2	2	2
9	2	2	1	2	2
10	2	2	2	2	2
11	2	3	2	3	2
12	2	1	1	1	1
13	1	1	1	1	1
14	3	3	3	3	3
15	1	1	1	1	1
16	1	1	1	1	1
17	2	3	2	2	2
18	1	1	2	1	1
19	2	2	2	2	2
20	2	2	3	2	3
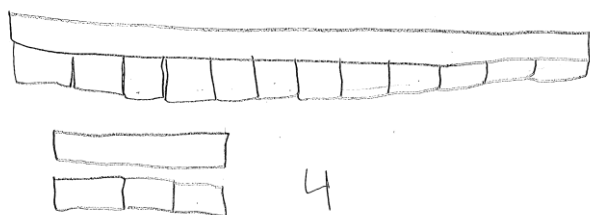
Figure 4. Student #1 response to Task 6

rubric. The remaining seven cases drew our attention to indicators that we either had not attended to while scoring, or that led to ambiguous scores.

In the cases of Students 1, 11, 17, and 20, there was some question as to whether students had drawn pictures fitting the given relations in order to demonstrate their reasoning (indicating Stage 3), or whether they had drawn the pictures in order to find a solution (indicating Stage 2). For example, in response to Task 6, Student #1 had drawn the picture shown in Figure 4.

On the basis of this response and similar responses to other tasks, R1 had assessed Student 1 as operating at Stage 3. However, analysis of the video-recorded clinical interview made it clear that the student had worked out his solution with the drawn bars: First, he drew the long bar with the 12 small bars lined up beneath; then he drew the medium bar, with the three small bars lined up beneath; then he determined how many sets of 3 there were among the 12 small bars. Although the student apparently reversed his multiplicative reasoning to produce the correct relationship, he seemed to rely on the drawn bars. Rather than assimilating the two given relationships into a three-level structure, the student seemed to assimilate the two given relationships into a two-level structure, then represented these two two-level structures in the drawing and used the drawing to coordinate those two structures. This way of operating fits Stage 2.

Looking back at the written responses, we see indicators that the student had operated in this way. First, note that the students' final, numerical response appears to the lower right of the drawings, indicating that he had produced it after the drawings. Second, note the pencil dots along the 12 small bars, indicating that the student counted them out, perhaps in groups of 3, after they were drawn. We can contrast these indicators with Student #20 who drew the picture shown in Figure 5.

Here, we note that the student had set up the drawing as a division problem from the start, but this was not a simple representation of a rote computation: The green ("g") and purple ("p") bars were drawn to scale. Moreover, the student had immediately translated the given relations into the essential relations between the number of green bars within each of the other bars, and he represented the result of their division with four purple bars, indicating that he had assimilated all three levels of units from the start. Finally, above the drawing he had written the following: "If the green bar fits into the orange bar 12 times and the green bar fits into the purple bar three times, you need to do 12 ÷ 3." This writing indicates that (1) the student determined the solution method before making the drawing and (2) the student saw the division as a *necessary* relationship ("you need to do") rather than a trick that might work.

Until R1 and R2 reviewed the video-recorded clinical interview with Student #20, they had missed these cues and had assumed the student was doing rote division. Distinctions in the written responses of Student #11 and Student #20 informed the indicators that we created for the rubric with respect to Tasks 5 and 6. Similar distinctions, as revealed by the clinical interviews, informed indicators across all seven tasks. R3 and R4 used the completed rubric to assess the 20 students' written responses to the seven tasks. These raters had no foreknowledge of our ratings or discussions in forming the rubric. We report on quantitative results regarding those assessments in the next section.

Validity and reliability of the instrument: Round 1

We tested the validity of the rater scores using the instrument by comparing the scores from R3 and R4 to the reconciled scores from R1 and R2, based on their assessments of students' actions in the videos. All three scores are reported in Table 3. Kendall's tau and Spearman's rho provide measures of association between the scores. For R3 and R4, respectively, tau was 0.89 ($p < .001$) and 0.81 ($p < .001$), and rho was 0.91 ($p < .001$) and 0.87 ($p < .001$). Inter-rater reliability for R3 and R4 was measured using Cohen's kappa: 0.32 (unweighted); 0.49 (weighted). Landis and Koch (1977) characterize these scores as indicating "fair" to "moderate" agreement.

On the second set of data, the first pair of raters (R1 and R2) achieved perfect agreement, with matching ratings for all 20 students. However, this data included only four students rated as Stage 2 and no students rated as Stage 3. As we began looking at the scores for the second pair of raters (R3 and R4), we realized that most of the disagreements were between Stages 2 and 3. So, when we met to further
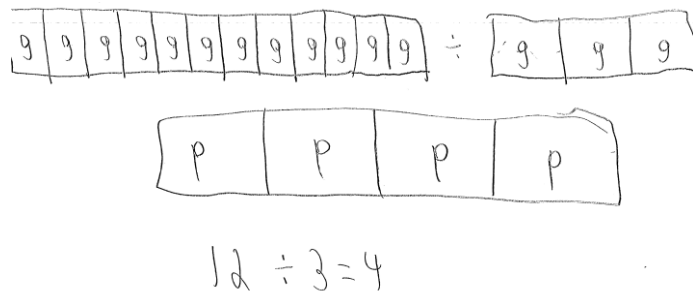

Figure 5. Student #20 response to Task 6

Table 3. Second pair of raters' assessments compared to final assessments from the first pair

Student#	Written R3	Written R4	Video Final
1	2	2	2
2	1	1	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	2	2
7	1	1	1
8	2	3	2
9	2	2	2
10	2	3	2
11	2	3	2
12	1	1	1
13	1	2	1
14	3	3	3
15	1	1	1
16	1	1	1
17	2	3	2
18	1	2	1
19	2	3	2
20	2	3	3

refine the rubric, we focused on creating indicators that would more clearly demark that distinction.

Rubric refinement

Following the first round of validity and reliability tests, R1 met with R3 and R4 to discuss discrepancies and review each indicator in the rubric for ambiguities. This resulted in a revised rubric (Appendix B) that would be used in the second round of testing with the third and fourth pairs of raters (R5-R8). In this section, we share our rationale for some of the revisions, especially regarding distinctions between Stages 2 and 3.

By definition, students operating at Stage 3 can assimilate three levels of units, whereas students operating at Stage 2 have to coordinate the third level of units through activity. Thus, a key indicator in distinguishing the two levels is to determine whether students used their drawings to coordinate the third level of units or whether those drawings were used to represent a structure they had already organized. Related indicators in the rubric required raters to infer whether students arrived at their solutions before or after they made their drawings. After looking back at cases in which either rater in the second pair (R3 and R4) had identified the indicator, we realized that we could specify more objective examples to guide rater inferences; we revised related indicators accordingly. For instance, the first Stage 2 indicator for Task 4 was revised to the following: "Students coordinate relations appropriately and with a drawing illustrating size relations, but writing indicates the drawing was the solution method (e.g., solution appears below the drawing, or erasures/corrections are present in the drawing)."

Other indicators seemed to be generating false positives for Stage 3. These included a second Stage 3 indicator for Task 4 ("students refer to necessary relationships") and the first Stage 3 indicator for Tasks 5 and 6 ("students reverse the multiplicative reasoning for both tasks, except possibly if there is indication that the student misread a task"). We edited the latter indicator by removing the exception, and we removed the former indicator altogether. The former indicator was designed as an attempt to account for the idea that, at Stage 3, students could immediately understand the third relationship as a logical consequence of the given relationships. However, in practice, it was ambiguous with indications possibly

including restatements of the given relations. The latter indicator was written so that raters would not exclude a student as Stage 3 for simply misreading one of the reverse tasks. However, we found that other Stage 3 indicators were adequate for identifying Stage 3 students.

Validity and reliability of the instrument: Round 2

We tested the validity and reliability of assessments based on the refined instrument by recruiting raters R5-R8. R5 and R6 (fifth and sixth authors) were mathematics education researchers familiar with the construct of units coordination. R7 and R8 were students in a teacher education program—one preparing to become a secondary school mathematics teacher (R7) and the other preparing to become an elementary school teacher (R8). Neither of these students had familiarity with the construct of units coordination beyond an article they read (Norton, Boyce, & Hatch, in press). The raters' scores are reported in Table 4, along with the reconciled scores (Video Final) from the first pair of raters.

Note that the preponderance of new raters' scores agree with the video-based "true" score. However, none of the scores from these raters agree with the video-based score for Student #9; in contrast, scores from R3 and R4 both agreed with this score (see Table 3). Student #9 responded to the final three tasks erroneously and with little supporting work: For Task 5, the student wrote "8÷2=6"; for Task 6, she wrote "12÷6=2"; and for Task 7, she labeled the medium and long bars "4" and "9", respectively, and simply wrote "2" as a final response. Contrasting the indicators identified by R3 and R4 using the old rubric and those identified by R5-R8 using the new rubric, we attribute the disparity in scores to the following refinements in the rubric.

1. For Tasks 5 and 6, the new rubric includes subtracting the numbers in the given relations as a Stage 1 indicator (Indicator 2), and although Student #9 wrote a division symbol, she appeared to subtract the numbers for Task 5.
2. Also for Tasks 5 and 6, the exception for misreading a task was removed from Stage 3 Indicator #1, so the new raters could not justify the Stage 3 indicator for Student #9, who did seem to misread Task 6.

Table 4. Third and fourth pairs of raters' assessments compared to final assessments from the first pair

Student#	Written R5	Written R6	Written R7	Written R8	Video Final
1	2	2	2	2	2
2	1	1	1	1	1
3	1	1	2	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	2	1	2	1	2
7	1	1	1	1	1
8	2	2	2	2	2
9	1	1	1	1	2
10	2	3	2	3	2
11	2	3	3	3	2
12	1	1	1	1	1
13	1	1	2	1	1
14	3	3	3	3	3
15	1	1	1	1	1
16	1	1	1	1	1
17	2	2	2	2	2
18	1	2	1	1	1
19	2	2	2	2	2
20	3	2	3	3	3

3. Finally, for Task 7, the new rubric included a Stage 1 indicator for when the students do not account for the leftover part, which Student #9 did not do.

Despite the refined indicators, video analysis indicates that Student #9 had interiorized two levels of units, as evidenced by the students' oral explanations of reverse multiplicative reasoning in Tasks 5 and 6 and attempts to account for the leftover part in Task 7. Thus, we view the raters' scores of written responses using the new rubric as false negatives for Stage 2 operating, as a consequence of this student's particular responses and the lack of supporting drawings or written explanations. Although we could continue to refine the rubric, our overall statistical measures demonstrate reliability and validity of the rubric in its current form (Appendix B).

As before, we measured inter-rater reliability by comparing each pair of raters' scores to each other using Cohen's kappa, and we measured validity by comparing each rater's scores to the reconciled scores from R1 and R2 (based on their assessments of students' actions in the videos) using Kendall's tau and Spearman's rho. For R5 and R6, kappa was 0.57 (unweighted) and 0.65 (weighted). For R7 and R8, kappa was 0.68 (unweighted) and 0.75 (weighted), indicating "substantial" agreement within each pair (Landis & Koch, 1977). Tau for the four raters (R5, R6, R7, R8) were 0.93, 0.68, 0.75, and 0.81, respectively; and Spearman rank-order correlations for the four raters were .93, .73, .78, and .84, respectively. These measures of association (see Table 5) represent strong relationships between the rater scores and the "true" scores (Cohen, 1992), providing strong evidence of the criterion-related validity for the rater scores based on the revised rubric.

DISCUSSION

The goal of this study was to develop an instrument for identifying stages of units coordination and to test the validity and reliability of assessments based on that instrument. Evidence presented here provides indication for the usefulness of the instrument—tasks and rubric—for assessing the levels of units with which middle school students' can operate. In particular, measures of validity and reliability among the third and fourth pairs of raters (R5-R8) indicate that teachers can use the instrument as well as researchers, with little training on theoretical underpinnings. Recent research suggests that teachers may need such an instrument for effective instruction. For example, with regard to algebraic reasoning, Hackenberg (2013b) demonstrated how productive engagement for students operating at Stage 1 might differ from productive engagement for students operating at Stages 2 and 3. As elaborated in our *Theoretical Framework*, additional studies implicate the role of units coordination across multiple domains of mathematics (e.g., Steffe & Olive, 2010; Tillema, 2013; Ulrich, 2012).

Researchers can use the instrument to measure growth in students' units coordination. To date, researchers have assessed students' stages of units coordinating using clinical interviews or small group teaching experiments (e.g., Hackenberg & Tillema, 2009; Norton & Boyce, in review), which are too time intensive to accommodate large populations. In contrast, written responses to the

Table 5. Measures of validity and reliability

Measure	R5	R6	R7	R8
Tau	0.93**	0.68*	0.75**	0.81**
Rho	0.93**	0.73**	0.78**	0.84**
Kappa (unweighted)	0.57**		0.68**	
Kappa (weighted)	0.65**		0.75**	

Note: *p < .01; **p < .001.

instrument's tasks can be gathered at once and scored relatively quickly using the rubric, making it possible to perform large-scale quantitative studies of students' stages. Such studies might establish the statistical significance of growth due to instructional intervention. They might also address hypotheses regarding units coordination and its relation to other aspects of mathematical development through measures of correlation.

Izsák and colleagues developed an instrument for assessing teachers' fractions knowledge before and after a professional development intervention (Izsák, Jacobsen, de Araujo, & Orrill, 2012). Using mixture Rasch modeling, they identified two subgroups of teachers based on their differentiated performance. The researchers surmised that the distinguishing feature of the higher-performing group was their ability to coordinate three levels of units. Likewise, Ellis (2007) found that a distinguishing feature for students' success in high school algebra was their ability to treat ratios as quantities—an ability that may depend upon coordinating three levels of units (Hackenberg, 2013a). Both hypotheses can be tested measuring correlation between student performance in these domains and their stages of operating with units.

Future research should include tests of convergent and divergent validity relative to other measures. For example, we might want to know whether stages of units coordination correlate strongly with Piaget's (1970a) operational stages. Middle school students are in a developmental range (around 11 years old) during which they should be transitioning from the concrete operational to formal operational stage. To what degree does this transition correlate with higher stages of units coordination? Measures of convergent and divergent validity could also differentiate the effects of general ability, whether framed as intelligence or general test-taking ability.

We have framed units coordination as a general structure for operating that crosses multiple mathematical domains. This framework fits Steffe's (2002) reorganization hypothesis concerning the ways students operate with whole numbers and fractions. More generally, it fits a Piagetian perspective on the nature of mathematics itself: "The whole of mathematics may be thought of in terms of the construction of structures" (Piaget, 1972, p. 70). Attending to students' structures for operating provides a sense of their mathematical power, including the kinds of concepts they might construct with the support of instruction. We intend the units coordination assessment instrument shared here as way for teachers and researchers to better understand students' mathematical development.

In closing, we note that the future studies suggested here might also address limitations in our own study. For example, we tested our instrument with a relatively small population in one geographic region. Using the instrument with new populations offers opportunity for further tests of reliability and validity—including tests of convergent and divergent validity in studies that use multiple instruments. We see our contribution as a first step by providing the research community with a means of assessing stages of units coordination among large populations of students.

ACKNOWLEDGEMENT

REFERENCES

Boyce, S., & Norton, A. (in review). Co-construction of fractions schemes and units coordinating structures. *Journal of Mathematical Behavior*.

Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal*, 171-186.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Ellis, A. B. (2007). The influence of reasoning with emergent quantities on students' generalizations. *Cognition & Instruction*, *25*, 439 – 478. doi:10.1080/07370000701632397

Hackenberg, A. J. (2007). Units coordination and the construction of improper fractions: A revision of the splitting hypothesis. *Journal of Mathematical Behavior, 26*, 27-47. doi:10.1016/j.jmathb.2007.03.002

Hackenberg, A. J. (2010). Students' reasoning with reversible multiplicative relationships. *Cognition and Instruction, 28*(4), 1-50. doi:10.1080/07370008.2010.511565

Hackenberg, A. J. (2013a) Musings on three epistemic algebraic students. In: Moore K. C., Steffe L. P. & Hatfield L. L. (eds.) Epistemic algebraic students: Emerging models of students' algebraic knowing. University of Wyoming: Laramie WY: 81–124.

Hackenberg, A. J. (2013b). The fractional knowledge and algebraic reasoning of students with the first multiplicative concept. *The Journal of Mathematical Behavior*, *32*(3), 538-563. doi:10.1016/j.jmathb.2013.06.007

Hackenberg, A. J., & Lee, M. Y. (2015). Relationships between students' fractional knowledge and equation writing. *Journal for Research in Mathematics Education, 46*(2), 196-243.

Hackenberg, A. J., & Tillema, E. S. (2009). Students' whole number multiplicative concepts: A critical constructive resource for fraction composition schemes. *Journal of Mathematical Behavior, 28*, 1-18. doi:10.1016/j.jmathb.2009.04.004

Hunting, R. P. (1983). Alan: A case study of knowledge of units and performance with fractions. *Journal for Research in Mathematics Education, 14*(3), 182-197. doi:10.2307/748381

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Izsák, A., Jacobsen, E., de Araujo, Z., & Orrill, C. H. (2012) Measuring mathematical knowledge for teaching fractions with drawn quantities. *Journal for Research in Mathematics Education, 43*(4). doi:10.5951/jresematheduc.43.4.0391

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. doi:10.2307/2529310

Norton, A., & Boyce, S. (in review). Provoking the Construction of a Structure for Coordinating *n+1* Levels of Units. *Journal of Mathematical Behavior*.

Norton, A., & Boyce, S. (2013). A cognitive core for common state standards. *Journal of Mathematical Behavior, 32*, 266-279. doi:10.1016/j.jmathb.2013.01.001

Norton, A., Boyce, S., Hatch, J. (in press). Coordinating units and constructing fractions at the Candy Depot. *Teaching Mathematics in the Middle School*.

Norton, A., Boyce, S., Ulrich, C., & Phillips. N. (2015). Students' units coordination activity: A cross-sectional analysis. *Journal of Mathematical Behavior, 39*, 51-66. doi:10.1016/j.jmathb.2015.05.001

Olive, J. & Çağlayan, G. (2008). Learner's difficulties with quantitative units in algebraic word problems and the teacher's interpretations of those difficulties. *International Journal of Science and Mathematics Education, 6*, 269–292. doi:10.1007/s10763-007-9107-6

Olive, J., & Steffe, L. P. (2002). The construction of an iterative fractional scheme: the case of Joe. *Journal of Mathematical Behavior, 20*(4), 413-437. doi:10.1016/S0732-3123(02)00086-X

Piaget, J. (1970a). *Genetic epistemology* (E. Duckworth, Trans.). New York: Norton.

Piaget, J. (1970b). *Structuralism.* New York: Basic Books.

Piaget, J. (1972). *The principles of genetic epistemology*. London: Routledge & Kegan Paul.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space* (Trans. F. J. Langdon & J. L. Lunzer). New York: Norton (Original work published in 1948).

Reynolds, A., & Wheatley, G. H. (1996). Elementary students' construction and coordination of units in an area setting. *Journal for Research in Mathematics Education*, 564-581. doi:10.2307/749848

Steffe, L. P. (1992). Schemes of action and operation involving composite units. *Learning and Individual Differences, 4*(3), 259-309. doi:10.1016/1041-6080(92)90005-Y

Steffe, L. P. (1994). Children's multiplying schemes. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 3-39). Albany, NY: State University of New York Press.

Steffe, L. P. (2007, April). *Problems in mathematics education*. Paper presented for the Senior Scholar Award of the Special Interest Group for Research in Mathematics Education (SIG-RME) at the annual conference of the American Educational Research Association in Chicago, Illinois.

Steffe, L. P., & Olive, J. (2010). *Children's fractional knowledge*. New York: Springer.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Erlbaum.

Steffe, L. P. & Ulrich, C. (2014). The constructivist teaching experiment. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education: SpringerReference*. Berlin, Germany: Springer. doi:10.1007/SpringerReference_313181

Tillema, E. S. (2013). A power meaning of multiplication: Three eighth graders' solutions of Cartesian product problems. *The Journal of Mathematical Behavior*, *32*(3), 331-352. doi:10.1016/j.jmathb.2013.03.006

Ulrich, C. (2012). Additive relationships and signed quantities. (Doctoral dissertation). Retrieved from http://www.libs.uga.edu/etd

❖ ❖ ❖

Appendix A: Units Coordination Tasks

Name:_____   Teacher:_____   Date:_____

Use the bars shown above to answer the following three questions:

1. How many times does the **Medium Yellow Bar** fit into the **Long Red Bar**?

    answer: ☐

2. How many times does the **Small Blue Bar** fit into the **Medium Yellow Bar**?

    answer: ☐

3. Use this information to figure out how many times the **Small Blue Bar** fits into the **Long Red Bar**?

    answer: ☐

    Use the space below to **draw a picture and explain** your answer.

1

Name:_____   **Teacher:**_____   **Date:**_____

Use the following information to answer questions about the bars shown above:

4. **Pretend** that the **Medium Purple Bar** fits into the **Long Orange Bar** *exactly* **2** times.

Pretend that the **Small Green Bar** fits into the **Medium Purple Bar** *exactly* **6** times.

Use this information to figure out how many times the **Small Green Bar** would fit into the **Long Orange Bar**?

answer:

Use the space below to **draw a picture and explain** your answer.

2

Name:_____   Teacher:_____   Date:_____

Use the following information to answer questions about the bars shown above:

5. Now **pretend** that the **Medium Purple Bar** fits into the **Long Orange Bar** *exactly* 2 times.

   Pretend that the **Small Green Bar** fits into the **Long Orange Bar** *exactly* 8 times.

   Use this information to figure out how many times the **Small Green Bar** would fit into the **Medium Purple Bar**?

   answer: [   ]

   Use the space below to **draw a picture and explain** your answer.

3

Name:_____   Teacher:_____   Date:_____

Use the following information to answer questions about the bars shown above:

6. Now **pretend** that the **Small Green Bar** fits into the **Long Orange Bar** *exactly* **12** times.

   Pretend that the **Small Green Bar** fits into the **Medium Purple Bar** *exactly* **3** times.

   Use this information to figure out how many times the **Medium Purple Bar** would fit into the **Long Orange Bar**?

   answer: [   ]

   Use the space below to **draw a picture and explain** your answer.

4

Name:_____   Teacher:_____   Date:_____

Use the following information to answer questions about the bars shown above:

7. Now **pretend** that the **Small Green Bar** fits into the **Long Orange Bar** *exactly* 9 times.

    Pretend that the **Small Green Bar** fits into the **Medium Purple Bar** *exactly* 4 times.

    How can you use this information to figure out how many times the **Medium Purple Bar** would fit into the **Long Orange Bar**?

    answer: [   ]

    Use the space below to **draw a picture and explain** your answer.

5

Appendix B: Revised Rubric for Assessing Units Coordination Structures

Tasks 1-3

	Students' Unit Structures	**Student Reasoning on Tasks 1-3**	**Written Indicators of Reasoning**
Stage 1	Students can take one level of units as given, and may coordinate two levels of units in activity.	Students physically or mentally iterate the small bar in the long bar (or segment the long bar with the short bar), relying on the appearance of the bars to determine how many times it would fit.	• Student responses to Tasks 1 and 2 are off by more than 1 from the correct relation (note that students at Stages 2 and 3 might represent the relations as unit fractions). • Student responses to Task 3 indicate that they estimated the number of times small bar fit into long bar (possibly further indicated by partitioning marks), rather than taking the product of responses to Tasks 1 & 2. • Students add their solutions to Tasks 1 and 2 to solve Task 3. • Students do not respond, or otherwise indicate they do not know.
Stage 2	Students can take two levels of units (a composite unit) as given, and may coordinate three levels of units in activity.	Students mentally iterate the medium bar within the long bar four times, with each iteration representing a 3 (i.e., 3, 6, 9, 12).	• Students use relational language (e.g., "every medium bar is 3 small bars"). • Student drawings incorporate the two relations determined in Tasks 1 and 2. • Student responses justify the use of multiplication.
Stage 3	Students can take three levels of units (a composite unit of composite units) as given, and can thus flexibly switch between two and three-level structures without reliance on figurative material.	Students use the given relations to determine that there are four 3s (small bars) in the long bar.	

Task 4

	Students' Unit Structures	**Student Reasoning on Task 4**	**Written Indicators of Reasoning**
Stage 1	Students can take one level of units as given, and may coordinate two levels of units in activity.	Students rely upon the appearance of the bars without using given relations.	• Students rely upon the appearance of the bars rather than using the given relations (e.g., partitioning/segmenting the given bars). • Students add or subtract the numbers given in the relations. • Students do not respond, or otherwise indicate they do not know.
Stage 2	Students can take two levels of units (a composite unit) as given, and may coordinate three levels of units in activity.	Students use the second given relation to form a composite unit that they can iterate through activity, by the number in the first given relation.	• Students coordinate relations appropriately and with a drawing illustrating size relations, but writing indicates the drawing was the solution method (e.g., solution appears below the drawing, or erasures/corrections are present in the drawing). • Student explanations and drawings appropriately refer to multiple two-level relations, but not a single three-level relation. • Student responses indicate use of multiplication without justification or illustration (possibly with a multiplication error).
Stage 3	Students can take three levels of units (a composite unit of composite units) as given, and can thus flexibly switch between two and three-level structures without reliance on figurative material.	Students take the first given relation as a composite unit that they mentally distribute across the units given in the second relation, thus justifying the use of multiplication.	• Student drawings are used to justify or illustrate appropriate solutions rather than to produce them (e.g., drawing is integrated with or appears below an explanation). • Student explanations and drawings refer to a single three-level relation, with appropriate size relations.

Tasks 5 and 6 [record indicators present for either task]

	Students' Unit Structures	Student Reasoning on Tasks 5 and 6	Written Indicators of Reasoning
Stage 1	Students can take one level of units as given, and may coordinate two levels of units in activity.	Students mentally iterate the short (medium) bar, imagining how many times it would fit into the medium (long) bar.	• Students rely upon the appearance of the bars rather than using the given relations (e.g., partitioning/segmenting the given bars). • Students add or subtract the numbers given in the relations. • Students do not respond, or otherwise indicate they do not know.
Stage 2	Students can take two levels of units (a composite unit) as given, and may coordinate three levels of units in activity.	Students use the two given two-level relations to generate representations with which to relate them figuratively.	• Student drawings or explanations indicate multiplicative reasoning but not reverse multiplicative reasoning (leading them to multiply instead of dividing, possibly because they misread the task). • Student responses indicate use of division, but without justification or supporting illustrations. • Students rely upon their drawings of the given relations to determine the unknown relation. • Student explanations and drawings appropriately refer to multiple two-level relations, but not a single three-level relation.
Stage 3	Students can take three levels of units (a composite unit of composite units) as given, and can thus flexibly switch between two and three-level structures without reliance on figurative material.	Students assimilate the two given two-level relations into a structure for coordinating all three levels.	• Students reverse their multiplicative reasoning for both tasks. • Student drawings are used to justify or illustrate appropriate solutions rather than to produce them. • Student explanations and drawings refer to a single three-level relation, with appropriate size relations. • Students use division in ways that are consistent with drawings and explanations.

Task 7

	Students' Unit Structures	Student Reasoning on Task 7	Written Indicators of Reasoning
Stage 1	Students can take one level of units as given, and may coordinate two levels of units in activity.	Students mentally iterate the medium bar, imagining how many times it would fit into the long bar.	• Students rely upon the appearance of the bars rather than using the given relations (e.g., partitioning/segmenting the given bars). • Students add or subtract the numbers given in the relations. • Students multiply the numbers in the given relations without any explanation. • Students do not respond, or otherwise indicate they do not know. • Students make no attempt to account for the leftover part.
Stage 2	Students can take two levels of units (a composite unit) as given, and may coordinate three levels of units in activity.	Students establish a composite unit of 4 and estimate how many of these fit into a length of 9.	• Students refer to fractional part as 1/9 rather than ¼. • Students respond with 2 and a remainder. • Student responses indicate use of division, but without justification or supporting illustrations. • Student drawings or explanations indicate multiplicative reasoning but not reverse multiplicative reasoning (leading them to multiply instead of dividing, possibly because they misread the task). • Student explanations and drawings appropriately refer to multiple two-level relations, but not a single three-level relation.
Stage 3	Students can take three levels of units (a composite unit of composite units) as given, and can thus flexibly switch between two and three-level structures without reliance on figurative material.	Students coordinate 9 as two 4s with one unit left over without losing the relationship between this unit and the others.	• Students appropriately account for the left over part with a fraction or a decimal (e.g., "2 ¼"). • Student drawings are used to justify or illustrate appropriate solutions rather than to produce them. • Student explanations and drawings refer to a single three-level relation, with appropriate size relations. • Students use division in ways that are consistent with drawings and explanations.