The Mathematics Educator
2013 Vol. 22, No. 2, 31–54

Validating a Written Instrument for Assessing Students' Fractions Schemes and Operations

Jesse L.M. Wilkins, Anderson Norton, and Steven J. Boyce

Previous research has documented schemes and operations that undergird students' understanding of fractions. This prior research was based, in large part, on small-group teaching experiments. However, written assessments are needed in order for teachers and researchers to assess students' ways of operating on a whole-class scale. In this study, scores from written instruments used to assess students' fractions schemes and operations were examined for validity and reliability. Scores from the written assessments were correlated with scores from clinical interviews of 33 sixth graders. Results suggest that the written instruments provide reliable and valid measures for assessing the partitive unit fraction scheme and the splitting operation. However, there is no such evidence regarding the partitive fraction scheme. Implications for teachers and researchers are considered, and possible explanations for the scores from the written assessments related to the partitive fraction scheme are discussed.

In 2010, Steffe and Olive published *Children's Fractional Knowledge*—a culmination of two decades of research on the schemes and operations that undergird students' understanding of fractions. Their work specifies a hypothetical learning trajectory (Simon, 1995) for students' progress from less sophisticated schemes, such as the part-whole scheme, toward superseding schemes that account for fractional sizes relative to the whole. The

Dr. Jesse Wilkins *is professor of mathematics education in the department of Teaching and Learning at Virginia Tech. His research focuses on quantitative literacy, in particular, students' development of their notions of probability and rational number, and their beliefs and attitudes associated with mathematics.*

Dr. Anderson Norton *is associate professor of mathematics education in the mathematics department at Virginia Tech. His research focuses on the epistemology and psychology of mathematics, especially in modeling the ways students construct mathematics.*

Steven Boyce *is a doctoral student in mathematics education in the mathematics department at Virginia Tech. His research focuses on the mathematical modeling of students' unit coordinating activity.*

part-whole scheme is a way of operating that supports students' recognition and production of proper fractions as *m* parts in the fraction out of *n* equal parts in the whole. For example, when students working with a part-whole scheme assimilate the task as a request to partition the stick into five equal parts and pull out three of them. However, this way of operating is very limited, especially in situations involving improper fractions, as assimilating 7/5 as 7 out of 5 makes little sense (Tzur, 1999).

Whether in terms of schemes, concepts, or subconstructs, several researchers have noted the limitations of part-whole reasoning (e.g., Behr, Harel, Post, & Lesh, 1992; Mack, 2001; Olive & Vomvoridi, 2006; Streefland, 1993) and its prevalence in U.S. curriculum and instruction (Li, Chen, & An, 2009; Pitkethly & Hunting, 1996; Watanabe, 2007). Thus, finding ways to support students' transcendence from part-whole reasoning is a significant problem for mathematics educators, especially in light of the new Common Core State Standards for Mathematics (CCSSM; 2010), which call for much more sophisticated conceptions as early as grade 4 (Norton & Boyce, in press). The learning trajectory specified by Steffe and Olive (2010) can provide this support, by helping teachers and researchers assess students' ways of operating with fractions and suggesting tasks that might provoke new ways of operating. However, these assessments have been based on time-intensive, longitudinal, small-group teaching experiments (Steffe & Thompson, 2000), which might prove impractical for teachers and researchers conducting studies with larger numbers of students.

The present study was motivated by the need for valid whole-class assessments of students' ways of operating. We have designed written instruments to do just that and to test, on a larger scale, many of the hypotheses arising out of the work begun by Steffe and Olive. In particular, we have tested whether the schemes identified through this work follow the specified hierarchy (Norton & Wilkins, 2009; Norton & Wilkins, 2012); we have tested whether operations, such as splitting, develop in the specified manner (Wilkins & Norton, 2011); and we have tested how such operations might be constructed (Norton & Wilkins, 2013). However, arguments for the validity of our assessments were limited to the face validity of items included in the written instrument, and the degree of fit between theorized and measured relationships between constructs.

The purpose of this paper is to further establish the validity of our assessments, by correlating results from written assessments with those from more time-intensive clinical interviews. We test this validity for three schemes/operations in particular: the partitive unit fraction scheme (PUFS), the more general partitive fraction scheme (PFS), and the splitting operation. We begin by describing these theorized constructs; then we share our methods for quantitatively testing validity. Results from the study hold implications for teachers and researchers who are interested in assessing students' ways of operating with fractions on a larger scale.

Theoretical Framework

Scheme Theory

Following von Glaserfeld's (1995) interpretation of the Piagetian concept of *scheme*, we define a cognitive scheme as an individual's established way of experiencing and operating in a particular situation. A cognitive scheme consists of three sequential components: a recognition template, operation(s), and an expected result (see Figure 1). In the sections below, we define these terms and exemplify them within the context of a part-whole scheme. We then elaborate on three operations (partitioning, iterating, and splitting), and two fractions schemes (PUFS and PFS) that pertain directly to the validation of our written assessments.

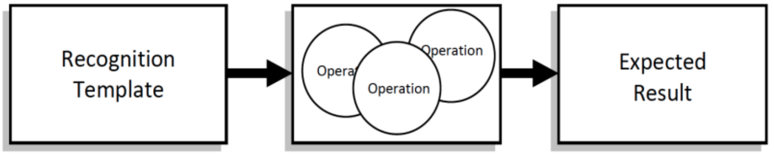

Figure 1. The three parts of a scheme.

The *recognition template* of a scheme is its assimilatory structure (Steffe, 2002). This structure is obtained from the sensory motor pattern awareness or mental imagery associated with an activity. Assimilation involves modification of perceptual input so that it fits into an individual's existing conceptual structures, and excludes characteristics of an individual's perception that are not part of those structures (von Glasersfeld, 1995). Thus, the recognition template functions as a guide for assimilation that triggers mental action from which satisfactory results are expected.

In the context of a cognitive scheme, an *operation* is an *interiorized mental action*. As mental actions, an individual's operations cannot be directly observed, but they might be inferred from accompanying verbal descriptions or physical actions (Tzur, 2007). Operations are abstractions of actions that "can be carried out through representation and not through actually being acted out" (Piaget, 1971, p. 14). This abstraction is what allows an operation to become coordinated with other operations (Piaget, 1971).

The *expected result* of a scheme is the anticipated product of operating according to that scheme (Steffe & Olive, 2010). Upon an individual's activation of a scheme, if an experiential result does not satisfy this expectation, then the individual experiences a *perturbation*. Resolving perturbations is the primary impetus for the *accommodation* of a scheme, which is a modification of one or more of the scheme's three components. Accommodations occur through the development of more powerful or efficient operations, or through a satisfactory assimilation of a novel situation into an existing scheme—a modification of the recognition template (von Glasersfeld, 1995).

To illustrate the components of a cognitive scheme, we describe a contextual situation, observable behavior, and hypothesized mental actions that we consider demonstrative of a student's part-whole fraction scheme. Two rectangular pieces of paper that have been previously segmented, one nearly white and one nearly black, are placed in a student's field of vision as depicted in Figure 2. A teacher says, "What fraction is the white bar out of the black bar?" Without hesitation, the student silently counts the number of parts in the white bar, moving her finger on each piece from left to right as she mouths the number words. After repeating this process with the black bar, she responds "three out of five."

For students who have constructed a part-whole fraction scheme, situations like the one illustrated in Figure 2 are likely to

Figure 2. Part-whole fraction task.

activate (or trigger) the scheme: The part and the whole are visibly partitioned into equally sized pieces, and the part-whole fraction scheme involves a whole number comparison of such pieces. Although the scheme includes partitioning, in this case that operation does not need to be carried out because the student can simply assimilate the existing partitions into the structure of the scheme, further supporting the likelihood that the situation would activate the scheme. Operating begins with the student disembedding the white rod from that black rod. That is, the white rod is mentally removed from (or taken "out of") the black rod without destroying the integrity of the black rod (Steffe & Olive, 2010). Now the two rods can be treated as collections of discrete uniform items, which leads to the use of a counting scheme and the forming of a fractional name (i.e., "three out of five") using the results of counting. The fractional name verbalized by the student fits the expected result of operating; hence, she does not experience perturbation, and no accommodation to the scheme is necessary.

Partitioning, Iterating, and Splitting Operations

The mental action of *partitioning* a continuous item might be inferred via an individual's placement of markings to indicate a separation of a whole into equal pieces. Partitioning involves the projection of a composite unit structure into a continuous whole[1]. By projection of a composite unit structure, we mean that a child uses her understanding of a composite number as a template. Before partitioning, the conception of the continuous item is of a single unit, and after partitioning, it is of a composite whole that contains equal parts. At first, the composite whole might not retain the entirety of the structure of the template (Steffe & Olive, 2010). For example, a child might conceive of a composite unit, *n*, as the result of iterating *n* ones, and the structure of such iteration might include a notion of nestedness of integers less than n. The child might project n into a continuous whole and use the structure to separate the whole into n equal parts that can subsequently be adjoined together, and this might result in a conception of a whole consisting of n $1/n$ parts. However, the result of this reuniting might not include a conception that $1/n$ is contained in $2/n$, is contained in $3/n$, … is contained in n/n.

The mental action of *iterating* requires the assimilation of an iterable unit—an object that can stand in for counting acts that have yet to be carried out (Steffe & Olive, 2010, p. 42). In the context of length, iteration might be indicated by the use of marks that correspond with the length of an iterable unit. Like partitioning, the structure of the result of iterating is only partially determined by the counting sequence by which it was constructed. For example, while the number n might stand in for the result of counting by ones n times, a child might not understand that to construct the number $n + 1$, she could simply apply n and then iterate one more. Similarly, the structure of the result of an iterating operation for fractions depends on the conception of the iterable unit and its relation with the whole. A child might iterate an object n times to make a whole, but attempting to iterate it another time to make an improper fraction, $(n + 1) / n$, might result in perturbation (Hackenberg, 2007; Olive & Steffe, 2001; Steffe & Olive, 2010; Tzur, 1999).

The *splitting*[2] operation is the simultaneous "composition of iterating and partitioning" (Steffe, 2004, p. 135). By simultaneous, we mean that the splitting operation is an interiorization of sequential partitioning and iterating. Previously the two operations might have been performed with concrete or mental re-presentations of specific concrete objects as they were modified and operated upon, but splitting enables an individual to instantly conceive of a single object in a more general and powerful way. Upon splitting an object, it has the structure of the result of partitioning the object into n parts, disembedding one of those parts, and iterating the part n times[3] (Steffe, 2002). The splitting operation enables the development of more powerful schemes because it allows this structure to become part of the recognition template for an existing scheme that previously involved sequential coordination of partitioning or iterating operations.

Figure 3 illustrates four tasks designed to elicit splitting from those students who have constructed the operation. Students who only coordinate partitioning and iterating sequentially would have difficulty assimilating the task into a scheme that would produce an appropriately sized piece. In other words, a student's splitting operation likely constitutes the necessary and sufficient condition for successful completion of each task: Each task is iterative in nature (calling for the production of a piece that is "*n* times as long/big") but requires partitioning to produce that piece; thus, to solve the tasks, students need to anticipate that partitioning and iterating act as inverse operations, the very criterion of splitting.

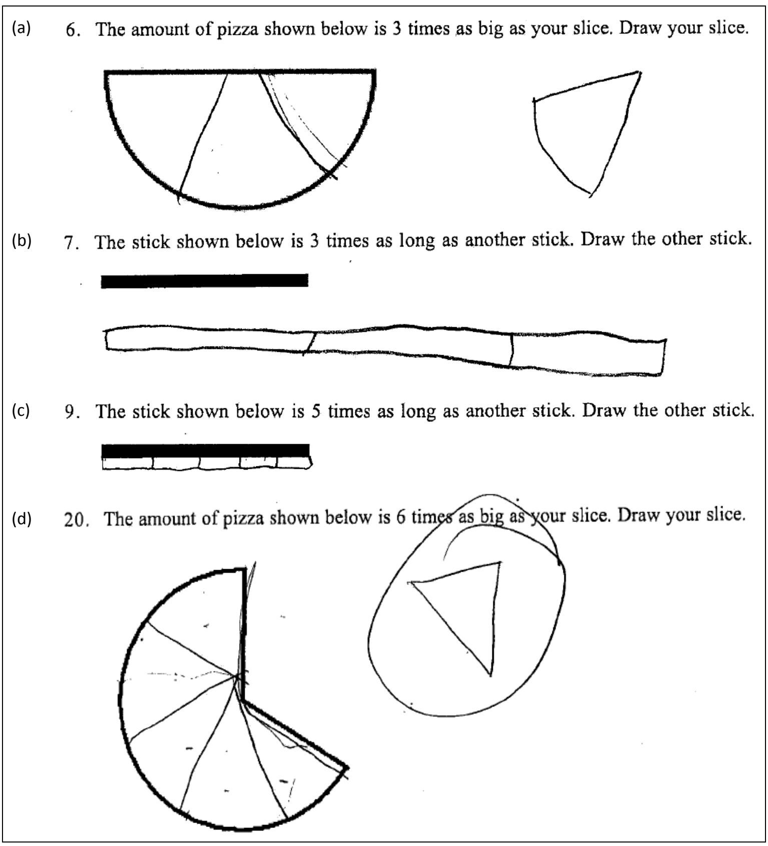

Figure 3. Tasks designed to trigger a splitting operation and examples of student responses.

Partitive Fraction Schemes

Steffe (2002) regards the partitive unit fraction scheme (PUFS) as the first genuine fraction scheme, because it includes fractional language in its recognition template and its operations include both iterating and partitioning (as well as disembedding). A student with a PUFS understands a unit fraction, such as $1/n$, as an amount that, when iterated n times, would result in a whole. When given a whole and instruction to make $1/n$ of the whole, a child assimilating the situation to a PUFS performs the following actions in sequence: First the child partitions the whole into n equal parts; next the child disembeds one of those parts so that it is separate from the whole without changing the whole; finally the child iterates that part n times to reproduce the whole. If the result is too small or too big, the child adjusts the size of the partition. If the result of iterating is satisfactory (that is, if it seems close enough in size to the given whole), then the fractional piece that was disembedded and iterated becomes "$1/n$."

The tasks illustrated in Figure 4 should trigger the PUFS, for those students who have constructed it. Students might successfully resolve the first two tasks in precisely the manner described above. However, students might also resolve those tasks using a part-whole scheme. Thus, it is important for teachers and researchers to attend to students' actions and infer from those actions the students' particular ways of operating. In the case of written assessments, we have to further infer those unobserved actions from the inscriptions that students make. The first two tasks in Figure 4 might provide further indication of students' ways of operating, to overcome this limitation. In particular, students operating with a PUFS can iterate the smaller piece within the whole to determine the fractional name of the piece. Students operating with a part-whole scheme do not perform such iterations.

The partitive fraction scheme (PFS) generalizes the PUFS to include construction and naming of fractions of the form m/n, where $1 < m < n$ (Steffe, 2002). If a student is given a task of producing m/n of a given whole, she might assimilate it into a PFS by first producing $1/n$ using a PUFS. This involves iterating the correct part 1, 2, …, m, …,n times. One might generalize the PUFS by reflecting on the construction of the composite fractions

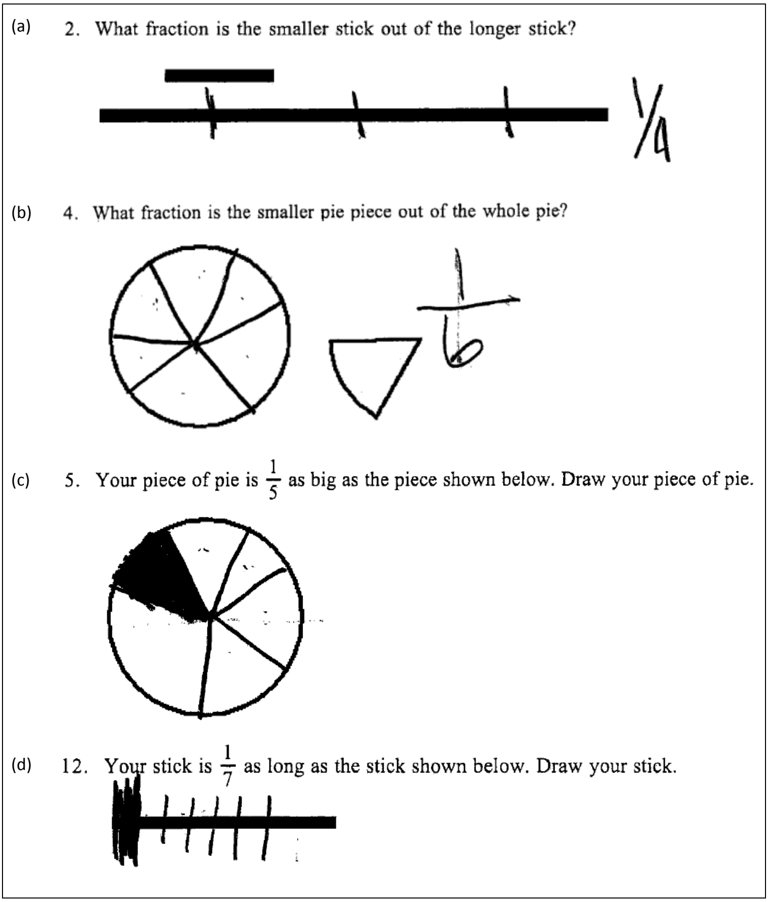

Figure 4. Tasks designed to trigger a partitive unit fraction scheme and examples of student responses

in the process of iterating, so that after iterating the correct piece n times, the intermediate products, including the fraction m/n, become assimilated as parts of the structure of the n/n whole. However, the structure of the m/n fraction is not necessarily m iterations of the fractional unit $1/n$, as this would be indicative of a more powerful *iterative fraction scheme* (IFS; Steffe, 2002). Rather, with the PFS, the $1/n$ piece is treated as a unit of 1, which does not maintain its 1-to-n relationship with the whole when being iterated (Gray, 1993; Steffe, 2002). As a result, students operating with a PFS (and not an IFS) might produce an improper fraction, such as 7/5, by iterating a 1/5 piece seven times, but they often name the result as “seven *sevenths*” (Tzur, 1999).

Figure 5 illustrates four tasks designed to elicit students’ use of a PFS. As with the PUFS tasks, the first two tasks could be solved using a part-whole scheme, but the language used in the tasks (e.g., “as long as”) was crafted to trigger schemes relating to size, which the part-whole scheme does not. Moreover, the final two tasks call on students to name fractions on the basis of a size comparison. Students who have constructed a PFS might make such comparisons based on their iterations of a unit fractional piece, *m* times and *n* times, respectively.

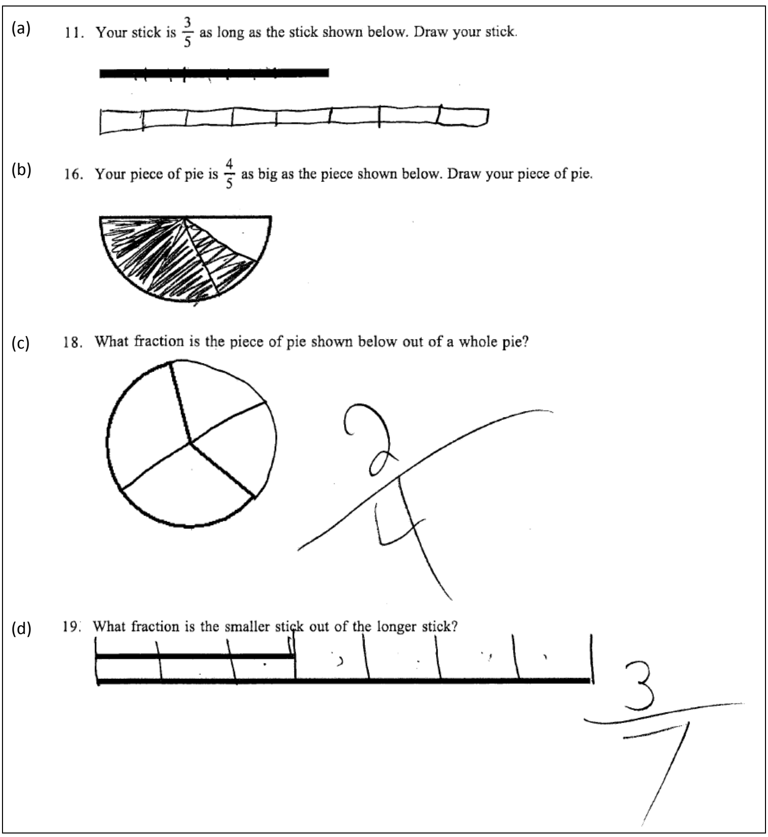

Figure 5. Tasks designed to trigger a partitive fraction scheme and examples of student responses.

Recent research has raised questions about the relationship between the construction of a splitting operation and the construction of a PFS (Norton & Wilkins, 2009; 2010). The PFS does not ostensibly require splitting[4], but written assessments of a PFS and splitting indicate that students generally construct a splitting operation before constructing a PFS (Norton & Wilkins, 2009, 2010; Wilkins & Norton, 2011). These findings highlight the difficulty of assessing the PFS. Students seem to have an established way of operating that includes the ability to solve some (but not all) tasks for which construction of a PFS is theoretically both necessary and sufficient.

Methods

Participants

Our study involved 66 students from three sixth-grade classrooms, all taught by the same teacher. The school is located in the rural Southeast, with 57% of its students eligible for free-and-reduced lunch programs. These 66 students participated in a written assessment of fractions schemes and operations; of these 66, 34 also agreed to participate in a clinical interview (Clement, 2000). Of these 34 students, after watching the video of the interviews, one student indicated that she did not feel well, and otherwise did not seem engaged in the tasks, making it difficult for the researchers to make meaningful inferences about her fractions understanding. For this reason, this student’s data were removed from further analysis, resulting in a sample of 33 students from the clinical interviews.

Written Assessment

We administered a 20-item assessment to the 66 participants. The 20 items included four items for the splitting operation (see Figure 3), the PUFS (see Figure 4), and the PFS (see Figure 5). The eight additional items were not analyzed in this study.

We designed individual items to provoke responses that might indicate a particular scheme or operation. In other words, each item provided students with a situation in which to enact particular ways of operating. The first and second authors independently rated student responses for each item. These ratings were based on all of the written work associated with the item and not solely on whether they had the "correct" answer. From this assessment, we inferred whether there was sufficient indication that the student had operated in a way that was consistent with a particular scheme or operation.

Following Norton and Wilkins (2009, p. 156; see also Wilkins & Norton, 2011), we scored responses to each item in the following way:

0: There was counterindication that the student could operate in a manner compatible with the theorized scheme or operation. Counter indication might include incorrect responses and markings that are incompatible with actions that would fit the scheme.

1: There was strong indication that the student operated in a manner compatible with the theorized scheme or operation. Indications might include correct responses, partitions and iterations

For item responses that represented some indication that a student operated in a manner compatible with the theorized scheme or operation, but nonetheless, did not show strong indication, we gave a score of .5. Initially, we also used scores of .4 or .6 to indicate a leaning, one way or the other. We used these scores to aid in the overall inferences about the schemes or operations. Inferences were based on overall scheme or operation scores created from the item scores. We discuss this procedure at the end of the section.

Here we provide some explanation of the types of student responses that would represent indication and counterindication of a particular scheme or operation and the scores associated with some of the responses. Referring to Figure 3, student response (a) shows a strong indication of a splitting operation (coded 1), whereas, in response (b) the student iterated the stick to create a stick three times as long as the given stick, representing strong counterindication that the student had a splitting operation available (coded 0). Student responses (c) and (d) represent weak counterindication (coded .4) and weak indication (coded .6), respectively. In response (c), the student partitioned the stick into five parts, but failed to identify the size of the stick. In response (d), based on the difference in the uneven partitions, it appears that the student divided the area into five parts and then realized that six parts were needed and drew in the squiggly partition. Although the resulting partitions were uneven, the student identified an appropriately sized piece—one that could be iterated six times to reproduce the given amount of pizza.

In Figure 4, student responses (a) and (b) provide strong indication of a PUFS (coded 1). In both examples the students appear to iterate the given piece within the given whole and, based on the resulting partitions, determine the fractional name for the piece. In response (c) the student begins by dividing the circle in half and then erases the line. Then the student fails to partition the area into the appropriate number of pieces. However, given that one part was identified, this represented weak counterindication of a PUFS (coded .4). In student response (d), the stick was first partitioned into eight pieces, but then the student seemed to catch the mistake and erased the last partition mark. The student did not adjust the other marks but identified one of the seven (unequal) parts. The lack of adjustment suggests a weak indication of the PUFS for this item (coded .6).

In Figure 5, student responses (b) and (d) provide strong indication of a PFS (coded 1). In response (b) the student partitioned the area into five parts, and then recognized that four of the pieces would create the desired piece of pie. However, based on the markings, it is not clear whether the student iterated a 1/5 piece (in accord with PFS) or whether the student used a part-whole scheme instead. In response (d), the student was able to partition the small stick into a piece that could be iterated to create the whole. Then the student was able to iterate the piece seven times within the whole and name the smaller stick. In response (a), the student partitioned the stick into five parts, but instead of iterating one of those parts three times, the student seemed to create a stick that was three more parts than the five-part whole. We considered this a weak counterindicaiton of a PFS (coded .4) because it is possible that the student interpreted the task as calling for a stick that is 3/5 longer than the given stick. Finally, student response (c) provides a counterindication of a PFS (coded 0) because the student named the fraction 2/4 despite the fact that the four pieces were clearly unequal and that the given piece was clearly bigger than 1/2.

For each set of four items representing a particular scheme or operation, the two raters independently summed the four individual item scores resulting in an overall raw score for each scheme between 0 and 4. These raw scores were then used to further infer whether the student had a particular scheme (coded as 1) or not (coded as 0). If a student's overall raw score for a given scheme or operation was greater than or equal to 3, it was inferred that the student's actions were consistent with the particular scheme or operation. If a student's overall score was less than or equal to 2, it was inferred that the student's actions were not consistent with the particular scheme or operation. For students whose overall score for a scheme was between 2 and 3, the individual raters considered the student's work on all four individual items and inferred from all of the work whether there was indication that the student had operated in a way that was compatible with the particular scheme or operation. For cases in which there was disagreement the two raters re-examined the cases together to come to a consensus.

Clinical Interviews

We conducted clinical interviews (Clement, 2000) with 34 students (one student's data were removed from the analysis). The clinical interviews were conducted by the three authors. The interviews involved eight tasks, with two tasks each for the PUFS, PFS, and splitting operation. The additional two tasks were not analyzed in this study.

Similar to the written assessments, we designed these tasks (see Table 1) to provoke student responses that might indicate a particular scheme or operation. Tasks were modified from Hackenberg and Lee (2012). The clinical interviews provided the interviewer the opportunity to ask clarifying questions and note actions that might not be apparent from a written assessment. For this reason, it was felt that the interviews would provide a more accurate assessment of students' fractions schemes and operations. Thus, the inferences from these interviews as to students' construction of a scheme or operation will serve as the criterion measure for evaluating the validity of the written assessments.

Each participant was interviewed by one of the three researchers for approximately 15-20 minutes, and interviews were videotaped and audio recorded. While the researcher followed the interview protocol (see Table 1), clarifying questions were asked when needed and notes were taken to document students' actions. Students' drawings were also collected as an additional artifact to support video analysis. Based on the interview, preliminary inferences of students' schemes and operations were noted by the interviewer.

Table 1.
Tasks and interview protocol for clinical interviews.

Splitting Tasks
(1) *[Provide 1 copy of the orange (8.5 cm) bar and a purple crayon]* Suppose the orange bar **is 5 times as long as** a purple bar. Can you draw the purple bar? *(a) If hesitant: Imagine what the purple bar would look like so that the orange bar would be* ***5 times as long as*** *the purple bar. Can you draw the purple bar?* *(b) After some guess: Can you show for sure that the orange bar is* ***5 times as long as*** *the purple bar?*
(2) *[Provide 1 copy of the red (4 cm) bar and a grey crayon]* Suppose the red bar **is 3 times as long as** a grey bar. Can you draw the grey bar?
PFS Tasks
(1) *[Place blue bar (28 cm) and black bar (20 cm) in front of the student]* What fraction is the black bar **of** the blue bar? *(a) If the student is hesitant to answer the question, or if there is other indication that the student doesn't know what the question is asking, then place the blue bar directly above the black bar (but not lined up) and say, "what fraction is the smaller bar of the larger bar?"* *(b) If student says a close fraction, e.g., ¾, ask how the student can be sure. If the student realizes that their initial guess is incorrect, ask what else it might be.*
(2) *[Provide 1 copy of the unpartitioned green bar and a brown crayon].* A brown bar is 4/7 **as long as** the green bar. Think about what the brown bar would look like. *(Allow manipulation of the green bar if hesitant).* Can you draw what you are thinking of? Can you show for sure that the brown bar is 4/7 **as long as** the green bar?
PUFS Tasks
(1) *[Place red bar (4 cm) and blue (28 cm) bar in front of the student]* What fraction is the red bar **of** the blue bar? *If the student is hesitant to answer the question, or if there is other indication that the student doesn't know what the question is asking, then place the red bar directly above the blue bar (but not lined up), and say, "what fraction is the smaller bar out of the larger bar?"*
(2) *[Provide 1 copy of the blue bar (28 cm) and a maroon crayon]* The maroon bar is 1/9 **as long as** the blue bar. Draw the maroon bar. Explain how you can show for sure that the maroon bar you drew is 1/9 as long as the blue bar.

Once all of the clinical interviews were completed, the three researchers met together and collectively analyzed each interview. Based on the viewing of the interviews, along with field notes and student work, we inferred whether there was sufficient indication that the student had operated in a way that was consistent with a particular scheme or operation. Similar to the coding that was used for the written assessments, if there was strong indication that the student operated in a manner compatible with the theorized scheme or operation, we coded them as having constructed the scheme or operation (coded as 1). If there was counterindication that the student could operate in a manner compatible with the theorized scheme or operation, we coded them as having not constructed the scheme or operation (coded as 0). For those students that showed some indication of operating in a manner compatible with the theorized scheme or operation, but nonetheless, did not show strong indication, or counter indication, we used scores of .4 or .6. These scores represented a leaning, in which a code of 0 or 1 could be assigned, respectively, in the case that a decision was needed, but at the same time, represented a lack of sufficient information to make a strong inference one way or the other.

We present the percentage of students that were determined to have constructed a PUFS, PFS, and a splitting operation in Table 2 by the different assessments.

Table 2.
Percentage of Students determined to have constructed a Splitting operation, a PUFS, and a PFS by assessment.

	Written Assessment	Clinical Interview
Splitting	48.5	54.6
PUFS	51.5	69.7
PFS	21.2	18.2

Note: These percentages are for the 33 students who were involved in both the written assessment and the clinical interview.

Evidence for the face validity of the scores associated with the written assessments has been presented previously (Norton & Wilkins, 2010; Wilkins & Norton, 2011). Here we present further evidence for the reliability and validity of the scores associated with the PUFS, PFS, and the splitting operation. We present an analysis of score reliability using measures of agreement (inter-rater reliability) and internal consistency (Cronbach's α). We then proceed to an analysis of score validity by analyzing the relationship between the scores from the written assessments with the scores from the clinical interviews (criterion-related validity). In this case, we are interested in how well we can predict students' clinical interview scores with the written assessments.

To assess the inter-rater reliability of the written assessments of the splitting operation, the PUFS and the PFS, we computed the kappa statistics (*K*; Cohen, 1960) for the scores from the 66 students from each rater (see Table 3). The kappa statistic for the splitting operation ($K = .88$, $p < .05$) and the PFS ($K = .82$, $p < .05$), represent "almost perfect" agreement (Landis & Koch, 1977, p. 165); and the kappa statistic for the PUFS ($K = .76$, $p < .05$) represents "substantial" agreement (Landis & Koch, 1977, p. 165). Overall the evidence suggests high inter-rater reliability for the scores. We discussed any differences in ratings and reconciled these differences to create one rating for each student. These resulting scores were used to compare to the findings from the clinical interviews.

Internal consistency is a measure of reliability that considers how well a set of items designed to measure a particular construct consistently measures the construct. To assess internal consistency, we calculated Cronbach's α (Cronbach, 1951) for the four items used to assess the PUFS, the PFS, and the spitting operation. Cronbach's α ranges from 0 to 1, and values around .70 and higher represent an acceptable level of reliability (Nunnally & Bernstein, 1994). Because the overall scores being evaluated were at the scheme and operation level, the individual items within a scheme or operation had different ratings by the two raters. For this reason we calculated Cronbach's α for each rater (see Table 3). The Cronbach's α for the PUFS and splitting operation from the two raters ($\alpha_{R1} = .69$ and $\alpha_{R2} = .72$) represent an acceptable level of internal consistency for the measures; however, the internal consistency for the PFS was not found to be acceptable ($\alpha_{R1} = .50$ and $\alpha_{R2} = .49$).

In order to assess the validity of the scores from the written assessments, we correlated these scores with those from the clinical interviews, to create validity coefficients, *r*. In the case of dichotomous scores, *r* represents a Phi coefficient. These validity coefficients provide a measure of criterion-related validity. Two validity coefficients were calculated for each scheme and operation. The first coefficient, r_1, was based on all 33 students. The second coefficient, r_2, was based only on those students for which the researchers made a strong inference for the students' schemes and operations based on the clinical interviews, that is, students coded with either a 1 or 0.

Table 3.
Measures of validity and reliability for the scores from the written assessments.

	K	α_{R1}	α_{R2}	r_1	r_2
Splitting	.88	.69	.72	$.52^{**}$	$.58^{**}$ (n = 32)
PUFS	.76	.69	.72	$.42^{*}$	$.53^{**}$ (n = 26)
PFS	.82	.50	.49	.14	.10 (n = 22)

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; r_1 = validity coefficient for scores on assessments with all 33 scores from interviews; r_2 = validity coefficient for the subset of scores on assessments with scores from interviews for which ratings were a 1 or 0.

Considering the PUFS (see Table 3), based on the data from all 33 students, PUFS scores from the written assessments correlated .42 ($p < .05$) with the scores from the clinical interviews. When considering only the students who were coded 0 or 1 (n = 26), the PUFS scores from the written assessments correlated .53 ($p < .01$) with the scores from the clinical interviews. These validity coefficients represent moderate to strong relationships between the scores from the written assessments and the scores from the clinical interviews (Cohen, 1992), providing evidence for the predictive validity for the scores from the written assessments associated with the PUFS.

Considering splitting (see Table 3), based on the data from all 33 students, the scores from the written assessments correlated .52 ($p < .01$) with the scores from the clinical interviews. Further, when considering only the students who were coded 0 or 1 (n = 32), the scores from the written assessments correlated .58 ($p < .01$) with the scores from the clinical interviews. These validity coefficients represent strong relationships between the scores from the written assessments and the scores from the clinical interviews (Cohen, 1992), providing strong evidence for the predictive validity for the scores from the written assessments associated with splitting.

Considering the PFS (see Table 3), based on the data from all 33 interviewed students, the scores from the written assessments correlated .14 ($p > .05$) with the scores from the clinical interviews. When considering only the students who were coded 0 or 1 ($n = 22$), the scores from the written assessments correlated .10 ($p > .05$) with the scores from the clinical interviews. Based on these validity coefficients the scores from the written assessments are not good predictors of the scores from the clinical interviews (Cohen, 1992), suggesting that the scores from the written assessments may not be valid measures of the PFS.

Conclusions and Implications

The goal of this study was to provide evidence for the validity of scores from written instruments used to assess students' fraction schemes and operations. Evidence presented here indicates that the written instruments provide reliable and valid measures for assessing the PUFS and splitting operation. However, we did not find sufficient evidence to suggest that they provide reliable and valid measures for assessing the PFS. Here, we consider two key implications from these conclusions. First, we consider the potential value of our results for teachers and researchers intending to use written instruments to assess students' schemes and operations. Then, we consider possible explanations for why our assessments using the PFS instrument did not yield positive results.

Scheme theory provides researchers with a framework for modeling students' mathematical ways of operating (von Glasersfeld & Steffe, 1991). However, assessing these ways of operating has relied upon intensive interactions with one or two students, in the form of clinical interviews or teaching experiments (Steffe & Thompson, 2000). With valid and reliable methods for assessing students' schemes and operations through the use of written instruments, researchers can more readily assess students' ways of operating at the classroom level. They can also use these instruments, as we have (Norton & Wilkins, 2010; Wilkins & Norton, 2011; Norton & Wilkins, 2012; Norton & Wilkins, 2013), to quantitatively test hypotheses about these ways of operating and the hypothetical learning trajectories related to them. Although written instruments provide relatively limited opportunity for observation from which to make inferences about students' mathematics, our results demonstrate moderate to strong correlations between assessments using the written instrument and assessments using clinical interviews, for the PUFS and splitting.

Teachers, too, can use the instruments to assess students' ways of operating with fractions. This would require teachers, first, to become familiar with the targeted schemes and operations, which have been described in a pair of teacher journal articles (Norton & McCloskey, 2008; McCloskey & Norton, 2009). However, teachers would not necessarily need to use the instruments in the manner we have described here—which involved use of two raters and inference across multiple tasks. Instead, they might use student responses to individual items as initial indicators of students' ways of operating with fractions, similar to the sample analysis described in the methods section. On the other hand, we warn that inferences based on single items provide an initial indication only, which might be used as a starting point for identifying students who have yet to develop more sophisticated fraction schemes. Moreover, teachers need to avoid the temptation to teach students how to respond to the items, lest they simply mask students' underlying ways of operating.

Having items and a reliable instrument for assessing the PUFS is particularly useful, given the importance of supporting students' transcendence from part-whole conceptions of fractions (Mack, 2001; Olive & Vomvoridi, 2006); likewise for the splitting operation, which plays a critical role in students' construction of advanced fractions schemes, such as the iterative fraction scheme (Steffe & Olive, 2010). However, the precise role of the PFS in students' development of fractions knowledge has proved problematic (Norton & Wilkins, 2010). In fact, teaching experiments on the construction of PFS have produced divergent characterizations of the scheme, so it is not surprising that our attempts to assess it with written instruments were unsuccessful.

Researchers (Hackenberg, 2007; Steffe, 2004) agree that generalizing a PUFS to a PFS involves iterating unit fractional parts to produce proper fractions (e.g., students can produce 4/9 from a 1/9 part by iterating the latter part four times). They also agree that the unit fractional part is iterated as if it were a unit of 1, rather than as a true unit fraction, which would maintain it's 1-to-n size relationship with the whole while being iterated (Tzur, 1999); as such, students still interpret the results of their iterations with their part-whole schemes. Hackenberg (2007) has argued that a partitive conception of, say, 3/5 derives from the part-whole conception alone, "not from it being a fraction that is three times one-fifth" (p. 30). However, Steffe (2004) claims that the partitive conceptions resulting from the PFS include considerations of size, wherein 7/10 refers to "the *length* of a stick produced by iterating a 1/10 stick seven times [emphasis added]" (p. 184). These contrasting perspectives raise questions about the nature of the PFS, its reversibility, and its relationship to conceptions of proper fractions as measures.

On the one hand, the operations of the PFS include nothing beyond those of the PUFS—*partitioning* a continuous whole into *n* equal parts, *disembedding* one of those parts from the partitioned whole, and *iterating* that part—except that, now, the part (treated as a unit of 1) is iterated *m* times to produce the proper fraction, as well as *n* times to reproduce the whole. On the other hand, at some point students begin to understand proper fractions as measures of size relative to the whole (as assessed by the tasks in Figure 5). Because these schemes exist only in the minds of researchers attempting to build explanatory models of students' mathematical activity, we need to decide whether it is useful to think of the PFS as a scheme at all, or whether some of what we attribute to the PFS is a simple extension of existing ways of operating, while other aspects of thinking attributed to the scheme involve more sophisticated ways of operating (e.g., the reversible partitive fraction scheme; Steffe & Olive, 2010).

References

Behr, M., Harel, G., & Post, T. A., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook for research on mathematics teaching and learning,* (pp. 296–333). New York: Macmillan Publishing Company.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547–589). Mahwah, NJ: Erlbaum.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and psychological measurement*, *20*(1), 37–46.

Common Core State Standards Initiative (2010). *Common Core State Standards for Mathematics.* Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Confrey, J. (1994). Splitting, similarity, and the rate of change: New approaches to multiplication and exponential functions. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 293–332). Albany, NY: State University of New York Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Gray, E. M. (1993, January). The transition from whole number to fraction. Paper presented at the International Study Group on the Rational Numbers of Arithmetic, University of Georgia, Athens, GA.

Hackenberg, A. J. (2007). Units coordination and the construction of improper fractions: A revision of the splitting hypothesis. *Journal of Mathematical Behavior, 26*(1), 27–47.

Hackenberg, A. J., & Lee, M. (2012). Fractional knowledge and equation writing. Manuscript submitted for publication.

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

Li, Y., Chen, X., & An, S. (2009). Conceptualizing and organizing content for teaching and learning in selected Chinese, Japanese, and U.S. mathematics textbooks: The case of fraction division. *ZDM, 41,* 809–826.

Mack, N. K. (2001). Building on informal knowledge through instruction in a complex content domain: Partitioning, units, and understanding multiplication of fractions. *Journal for Research in Mathematics Education, 32*(3), 267–296.

McCloskey, A., & Norton, A. (2009). Modeling students' mathematics using Steffe's advanced fractions schemes. *Mathematics Teaching in the Middle School, 15*(1), 44–56.

Norton, A. & Boyce, S. (in press). A cognitive core for common state standards. *Journal of Mathematical Behavior*.

Norton, A., & McCloskey, A. (2008). Teaching experiments and professional development. *Journal of Mathematics Teacher Education, 11*(4), 285–305.

Norton, A. & Wilkins, J. L. M. (2009). A quantitative analysis of children's splitting operations and fraction schemes. *Journal of Mathematical Behavior*, *28*(2/3), 150–161.

Norton, A., & Wilkins, J. L. M. (2010). Students' partitive reasoning. *Journal of Mathematical Behavior*, *29*(4), 181–194.

Norton, A., & Wilkins, J. L. M. (2012). The splitting group. *Journal for Research in Mathematics Education, 43*(5), 557–583.

Norton, A., & Wilkins, J. L. M. (2013). Supporting students' constructions of the splitting operation. *Cognition and Instruction*, *31*(1), 2–28

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Olive, J., & Steffe, L. P. (2001). The construction of an iterative fractional scheme: the case of Joe. *Journal of Mathematical Behavior, 20*(4), 413–437.

Olive, J., & Vomvoridi, E. (2006). Making sense of instruction on fractions when a student lacks necessary fractional schemes: The case of Tim. *Journal of Mathematical Behavior 25*(1), 18–45.

Piaget, J. (1971). *Psychology and epistemology* (A. Rosin, Trans.). New York: Grossman.

Pitkethly, A., & Hunting, R. (1996). A review of recent research in the area of initial fraction concepts. *Educational Studies of Mathematics, 30*(1), 5–38.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.

Steffe, L. P. (2002). A new hypothesis concerning children's fractional knowledge. *Journal of Mathematical Behavior*, *20*(3), 267–307.

Steffe, L. P. (2004). On the construction of learning trajectories of children: The case of commensurate fractions. *Mathematical Thinking and Learning, 6*(2), 129–162.

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

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

Streefland, L. (1993). Fractions: A realistic approach. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 289–326). Hillsdale, NJ: Erlbaum.

Tzur, R. (1999). An integrated study of children's construction of improper fractions and the teacher's role in promoting that learning. *Journal for Research in Mathematics Education, 30*(4), 390–416.

Tzur, R. (2007). Fine grain assessment of students' mathematical understanding: participatory and anticipatory stages in learning a new mathematical conception. *Educational Studies in Mathematics*, *66,* 273–291.

Watanabe, T. (2007). Initial treatment of fractions in Japanese textbooks. *Focus on Learning Problems in Mathematics, 29*(2), 41–60.

Wilkins, J. L. M., & Norton, A. (2011). The splitting loope. *Journal for Research in Mathematics Education, 42*(4), 386–416.

von Glasersfeld, E. (1995). A constructivist approach to teaching. In L. P. Steffe, & J. Gale (Eds.), *Constructivism in education* (pp. 3–15). Hillsdale, NJ: Erlbaum.

von Glasersfeld, E., & Steffe, L. P. (1991). Conceptual models in educational research and practice. *Journal of Educational Thought, 25*(2), 91–103.

---

1 Steffe and Olive (2010, chapter 4) refers to the distinctions between the two types of partitioning mentioned here as the third and fourth levels of fragmenting. At earlier, non-interiorized levels of fragmenting, children use strategies that are more empirical in nature.

2 The definition of splitting we use here is more restrictive than Confrey’s (1994) definition in which splitting refers to a possibly recursive operation that results in the production of many parts simultaneously from a whole.

3 These are the defining operations of an *equipartitioning* scheme, which is closely related to the PUFS (see Steffe, 2002).

4 A *reversible partitive fraction* scheme is theorized to be the least powerful scheme that requires splitting (Hackenberg, 2007; Norton & Wilkins, 2010; Steffe & Olive, 2010).