ORIGINAL ARTICLE

Development and application of a two-tier diagnostic instrument to assess middle-years students' proportional reasoning

Annette Hilton · Geoff Hilton · Shelley Dole · Merrilyn Goos

Abstract Proportional reasoning involves the use of ratios in the comparison of quantities. While it is a key aspect of numeracy, particularly in the middle years of schooling, students do not always develop proportional reasoning naturally. Research suggests that many students do not apply proportional methods appropriately and that they often erroneously apply both multiplicative and additive thinking. Further, students cannot always distinguish non-proportional situations from those that are proportional. Understanding the situations in which students mistakenly use additive or multiplicative thinking and the nature of the proportional reasoning that students apply to different problem types is important for teachers seeking to support their students to develop proportional reasoning in the classroom. This paper describes the development and use of a two-tier diagnostic instrument to identify situations in which students could and could not apply proportional reasoning and the types of reasoning they used. It presents data from an Australian study involving over 2000 middle-years students (Years 5 to 9) as a means of illustrating the use of the instrument for diagnosing students' reasoning in different situations. The findings showed that the instrument was useful for identifying problem types in which students of different ages were able to apply correct reasoning. It also allowed identification of the types of incorrect reasoning used by students. The

A. Hilton (✉)
Aarhus University, Copenhagen, Denmark
e-mail: anhi@dpu.dk

A. Hilton · G. Hilton · S. Dole · M. Goos
The University of Queensland, Brisbane, Australia

G. Hilton
e-mail: g.hilton@uq.edu.au

S. Dole
e-mail: s.dole@uq.edu.au

M. Goos
e-mail: m.goos@uq.edu.au

paper also describes useful applications of the instrument, including its use as a diagnostic instrument by classroom teachers and its use in the design of classroom activities included in teacher professional learning workshops.

Keywords Proportional reasoning · Middle-school mathematics · Diagnostic testing · Two-tier testing

Introduction

Proportional reasoning is mathematical reasoning involving a sense of co-variation and multiple comparisons (Lesh et al. 1988). Additive reasoning involves considering the sums or differences in quantities, whereas proportional reasoning requires multiplicative and relational thinking and involves the use of ratios in the comparison of quantities (Bright et al. 2003). These skills do not always develop naturally in students (Sowder et al. 1998). Lesh et al. (1988) described proportional reasoning as the cornerstone of higher-level areas of mathematics, such as algebra, and the capstone of elementary concepts, such as arithmetic, number, and measurement. The importance of proportional reasoning goes beyond the mathematics classroom; its importance in other areas of study, for example, science, economics, and demographics, has been widely recognised (Akatugba and Wallace 2009; Boyer et al. 2008). Perhaps more importantly, proportional reasoning is an essential aspect of everyday applications of numeracy. According to Ahl et al. (1992), proportional reasoning is a "pervasive activity that transcends topical barriers in adult life" (p. 81) and, yet, it has been estimated that more than half of the adult population are not proportional thinkers (Lamon 2005).

Until recently, little was known about students' development or application of proportional reasoning (Lamon 2005). Establishing the types and accuracy of reasoning used by students in situations of ratio and proportion can assist teachers in selecting appropriate teaching strategies and learning activities to target students' learning needs and strengthen their ability to reason proportionally (Misailidou and Williams 2003). The authors of this paper are conducting a large study that aims to enhance the teaching and learning of proportional reasoning in Australian middle-years classrooms. The purpose of this paper is to describe the development and use of a diagnostic instrument to identify the types of reasoning used by students in a range of proportional and non-proportional situations. It presents results from the administration of the instrument at the beginning of the study to illustrate its use in diagnosing students' reasoning in proportional and non-proportional situations. Finally, the application of the findings to teacher professional learning workshops, conducted with the teachers whose students completed the diagnostic instrument, is described.

Background

A review of the literature revealed varied descriptions of proportional reasoning problem types and the circumstances in which students typically use reasoning strategies correctly and incorrectly.

Problem types

There are two main categories of proportional problem types generally described in the literature (Cramer et al. 1993; Park et al. 2010). The first of these involves numerical comparison in which four values are given and the relation between them is to be determined. The second category involves missing-value problems, which provide three of four values and the proportional relationship between the four values. Within these categories, Lamon (1993) identified four semantic proportional problem types: well-chunked or well-known measures (relationships between two measures that result in a rate, which is itself a commonly used entity, e.g., distance/time=speed); part-part-whole (ratio problems in which two complementary parts are compared with each other or the whole); associated sets (rate situations in which the relationship between quantities is defined within the question, e.g., birthday cake pieces and children at a party); and stretchers and shrinkers or growth problems (situations that involve scaling up or down).

Lesh et al. (1988) criticised early attempts to assess students' proportional reasoning ability because of their limited focus on solving missing-value problems. They argued that such problems lend themselves to algorithmic approaches and do not necessarily provide students with an opportunity to apply proportional reasoning in a wide range of situations, an argument that aligns with Lamon's (1993) findings. Lesh et al. (1988) examined middle-school curriculum to identify the topics in which proportion-related problems arise. They described seven problem types and argued that, other than missing-value and comparison problems, the remaining five problem types are neglected in textbooks, instruction, and research. The neglected problem types included problems involving translations between or within representational modes, proportions involving unit labels, and mean-value problems.

Common proportional reasoning difficulties

Different problem types elicit different forms of reasoning (Langrell and Swafford 2000) and students experience varying degrees of difficulty with different problem types. For example, Lamon (1993) reported that students experience most success with problems involving associated sets and find growth problems the most challenging. Cramer et al. (1993) reported similar findings regarding growth problems and also found that problems involving non-integer relationships are more difficult than those involving integer relationships, even though the relationships and underlying concepts are no different.

The "conceptual stumbling blocks" critical to primary school students' proportional reasoning include understanding part-whole relationships; composite units, for example, those applied to rates; representation-related problems; and measurement-related abilities (Lesh et al. 1988, p. 9). Students often find it difficult to identify the appropriate reasoning to use in a particular situation. For example, Van Dooren et al. (2005) suggested that students have a tendency to overgeneralise the range of situations in which proportional thinking is applicable. They identified three non-proportional problem types in which students often rely on proportional reasoning when it is not required: constant, linear, and additive situations. Other research associated with proportional situations and the errors commonly made by students has identified similar problem areas:

- an inability to discern when to use proportional reasoning and difficulty in identifying multiplicative or relative relationships (Van De Walle et al. 2010);
- a tendency to approach proportional situations additively instead of multiplicatively (Cramer and Post 1993; Lamon 1993; Lesh et al. 1988; Misailidou and Williams 2003; Tourniaire and Pulos 1985);
- the use of multiplicative approaches unnecessarily, for example, when additive strategies should be used (Van Dooren et al. 2010);
- ignoring of some data, for example, attending only to numerators in an equation (Lesh et al. 1988);
- relating factors in a proportion in a qualitative manner, rather than applying quantitative strategies (Lesh et al. 1988);
- inappropriate use of algorithms, such as cross multiplication (Lesh et al. 1988; Nabors 2003); and
- incorrect build-up/pattern building (Lamon 1993; Misailidou and Williams 2003).

The problem types and common difficulties described above were important considerations in the design of the diagnostic instrument described in this paper.

Assessing proportional reasoning skills

Research shows that the understanding of proportionality develops slowly over several years and that teachers can support its development (Cramer and Post 1993; Howe et al. 2010; Lamon 2005). Several researchers have highlighted the importance of establishing children's proportional reasoning skills and the difficulties they experience with different problem types as a useful first step for teachers when selecting instructional activities to support students to develop appropriate additive and multiplicative reasoning strategies (e.g., Bright et al. 2003; Misailidou and Williams 2003; Van Dooren et al. 2005). This suggests that a diagnostic instrument would be useful for teachers to inform their choice of instructional activities to target their students' particular proportional reasoning learning needs.

Methods that can be used to determine students' understanding include interviews, pen-and-paper tests (open-ended or multiple-choice items), and concept mapping (Tüysüz 2009). Many such methods have been employed by researchers seeking to investigate students' proportional thinking over the years. For example, Lamon (1993) pretested 138 Year 6 students using an eight-item pen-and-paper test, which was followed by the use of clinical interviews with 24 of the same students, selected on the basis of their performance on the test. The students' responses were analysed according to whether they used absolute or relative thinking, additive or multiplicative strategies, and qualitative or quantitative proportional reasoning, and the sophistication of their strategies. Langrell and Swafford (2000) also used Lamon's categories to interview 16 students from Years 5 to 8 to investigate the strategies used by students to solve different problem types. The Rational Number Project tested students in Years 7 and 8 using missing-value problems, numerical comparison problems, and problems involving qualitative relationships (Cramer et al. 1993). Modestou and Gagatsis (2007; 2010) emphasised the limitations of focusing on missing-value and comparison problems. They argued that proportional reasoning involves an understanding of verbal and

analogical problems and the ability to discern non-proportional situations and to distinguish between direct and inverse proportions. Similar arguments have come from Sowder et al. (1998), who suggested that proportional reasoning reaches far beyond problems involving missing values and comparison to include understandings associated with equivalence, algebraic relationships, and transformations.

Dole et al. (2007) assessed the proportional reasoning of 800 middle-years students with a 16-item pen-and-paper test. Students were required to show all working and to explain their thinking for each item. All of the items focused on proportional reasoning. The item types included missing-value questions, rate problems, and relative thinking questions. There were also scale and inverse proportion questions. None involved non-proportional situations. The items were coded using a three-level code, which indicated whether the response was correct or incorrect, the nature of the response, and the thinking used by students.

Bright et al. (2003) emphasised the importance of providing situations in which students can correctly and incorrectly apply both multiplicative and additive thinking. They designed an instrument to assess Year 8 and 9 students' proportional reasoning that involved the use of four multiple-choice questions and one constructed response item. Three of the four multiple-choice items they used required proportional reasoning, and one required additive thinking. Although the authors acknowledged that having students share their reasoning on the items would be an effective classroom strategy, none of the multiple-choice items required students to justify or explain their answer on the test.

According to Wiggins and McTighe (1998), assessment must require students to explain or defend their answer. While open-response test items and interviews are powerful methods of establishing students' understanding, they can be impractical in many situations, for example, in research situations where large numbers of students are involved. Even open-response items of the type used by Dole et al. (2007) are difficult and time-consuming to code and analyse when conducting research with large numbers of students. Tamir (1989) found that the need for students to provide justifications when answering multiple-choice questions is a sensitive and effective means of assessing students' understanding, and argued that such an approach addresses some limitations commonly associated with traditional multiple-choice questions. To account for these arguments, Treagust (1995) recommended the use of two-tier multiple-choice instruments as an appropriate alternative to individual open-response questions or interviews as a means of obtaining information about the reasoning of groups of students. The first tier of a two-tier item consists of a multiple-choice question, with two to four choices. This may be a statement to which the student responds by selecting *true* or *false*, or it may be a more traditional multiple-choice question. The second tier usually requires students to choose from four reasons to justify or explain their answer to the first-tier question. The design of the options in the second tier is based on research findings, or on students' answers to free-response test items. The options consist of the designated correct response and identified misconceptions or reasoning errors (see Treagust 2006). In the case of the instrument described in this paper, both approaches were utilised. To be considered correct, students must answer both levels of the question correctly. The data generated are generally used to provide an insight into the common reasoning strategies employed by classes or year-level cohorts.

For a number of years, science education researchers have designed and success-fully used two-tier multiple-choice instruments to diagnose students' understanding and to identify alternative conceptions in a number of topics, as a means of informing pedagogical strategies (Chandrasegaran et al. 2008; Haslam and Treagust 1987; Özmen 2008; Tan and Treagust 1999; Treagust 1995, 2006; Tüysüz 2009). O'Keefe and O'Donoghue (2011) used a two-tiered instrument based on Treagust's work to investigate the effectiveness of an intervention strategy associated with fractions instruction in lower secondary mathematics classrooms; however, they did not provide examples of the items used. To date, such instruments have not been widely used for investigating students' understanding in mathematics.

Method

This section begins with an outline of the main considerations in designing the diag-nostic instrument and the administration procedure for the pilot of the instrument. This is followed by a description of the administration of the instrument in the main study.

Pilot instrument design

The design of many two-tier items used in science education research has been based on findings about students' common alternative conceptions or the errors and diffi-culties they commonly encounter in a particular topic, concept, or reasoning situation. The item types on the instrument described here were chosen to align with the arguments in the literature for the use of a range of item types and the need to ensure that both proportional and non-proportional items are included (Lesh et al. 1988; Van Dooren et al. 2005). The problem types are shown in Table 1 (examples have been included for clarification).

The first tiers of all problems were presented as a scenario followed by a statement to which students responded *True* or *False*. The second-tier responses of each item were used to identify students' correct and incorrect application of additive and multiplicative thinking, or their ability to distinguish situations of proportion from non-proportion; their recognition of absolute and relative situations; or their ability to interpret and transform representations of proportional situations. Each option gave an insight into the strategies the students used. The responses in the second tier of each item were based on the common strategies used by students in situations of proportion and non-proportion reported in the research literature and described earlier (e.g., Lamon 1993; Lesh et al. 1988; Van De Walle et al. 2010; Van Dooren et al. 2010). In addition, the findings of the study by Dole et al. (2007), which involved some members of the current research team, were used. Nine of the items on the open-response *Keeping it in Proportion* (KIIP) test used in the Dole et al. (2007) study were adapted for use on the two-tier instrument. The remaining three items were a constant non-proportional item, an additive non-proportional item, and a two-dimensional scale item. These were included to ensure a wider range of problem types and to address arguments in the literature that both non-proportional and proportional situations should be assessed (e.g., Lamon 2005; Van Dooren et al. 2005). The response patterns of the 800 students to items on the KIIP test were used

Table 1 Problem types and examples

Problem type	Example/description
Non-proportional: Constant	A group of children sings a song. If we double the number of children, how long will it take to sing the song?
Non-proportional: Additive	Two children run around a track at the same pace. One child starts two laps before the other. How far will the second have run when the first completes a given number of laps?
Missing value; associated sets; part-part-whole	If my recipe requires 10 cups of flour for 4 cups of sugar, how much flour will I need if I use 6 cups of sugar?
One- or Two-dimensional scale	If I double the length and width of a diagram, what will happen to the area of the diagram?
Familiar (well-chunked) rate; Inverse relationships	If a child runs at a certain pace and it takes a given amount of time, what will happen to the time if he runs at twice the pace?
Rate; translation of representations	Linking a verbal problem about speed and time with a graphical representation of the situation described
Relative thinking; associated sets	Given data about the number of children who choose an activity and the total children in each of two classes, identify which activity was relatively the most popular
Inverse proportion	If it takes 6 people 3 days to paint my fence, how many will it take to paint the fence in 2 days?

in the design of the second-tier alternatives of nine of the items. The second tier options for the other three items were designed using the findings of researchers, such as Van Dooren et al. (2005), on whose questions our non-proportional items were modelled. To illustrate the structure of the two-tier items used, an example is provided.

In the item, students were asked to compare two washing powders and to respond *True* or *False* to the statement "Powder A is the better value." The students were provided with a graphic that portrayed the following information:

Powder A comes in a 1 kg container that costs $4 for 20 loads of washing.
Powder B comes in a 1.5 kg container that costs $6.50 for 30 loads of washing.

The first-tier response choices were

T (true)
F (false)

The second-tier response choices were

A. Washing powder A costs the least.
B. Washing powder B costs a little bit more but you get ten more loads of washing.
C. The cost per load of washing is less.
D. Both washing powders are the same value.

This problem requires students to use relative thinking in a situation involving an associated set. Option 1 indicates that the student has used absolute thinking, comparing only price. Option 2 shows that the student has considered only the number of loads and not the price, again employing absolute thinking. Option 3 is the most accurate response

and shows that the student has used relative thinking. Option 4 suggests that the student has considered the mass and number of loads but not the price in his or her reasoning.

In order to avoid complications due to variation in problem structure and complexity of calculations, it was decided that the first tier of each item would be written in the form of a true-false statement. Because the instrument was intended for middle-years students (who in Australia are students in Years 5 to 9), the language used in the items was kept as simple as possible. The order of the second-tier responses was such that no item had TA as a correct response pair. This was done to ensure that students who guessed an answer by selecting the first option in each tier, or who did not engage with the task, were not credited with a correct response. The mathematical calculations were of low complexity in the majority of cases so as to focus on students' application of proportional thinking, rather than complicate the question with more challenging arithmetic. This said, both integer and non-integer proportional relationships were included to address the findings of Van Dooren et al. (2010) that the nature of the numbers included in word problems can influence students' reasoning. To prevent order effects, the test did not begin with a proportional question. This was done to avoid creating an expectation in the students that all items required multiplicative thinking, an approach used by Van Dooren et al. (2005). In addition, some second-tier responses were designed to determine whether students used qualitative or quantitative reasoning (see Lamon 1993; Langrell and Swafford 2000; Lesh et al. 1988).

Administration of the pilot instrument

Prior to administering the instrument to the pilot group, the items were reviewed and revised by members of the research team (who have extensive mathematics teaching experience in primary and/or secondary schools) and a sample of teachers. The pilot instrument was administered to 140 Year 5 and 6 students (11–12 years old) in composite classes in two primary schools, neither of which was participating in the main study. Three classes of students from each school completed the two-tiered instrument. The school principals were briefed on the purpose and administration procedure for the instrument and the teachers were provided with written instruction regarding its administration. The classroom teachers administered the test and were asked to read the items to the students if they felt this was necessary. There was no time limit imposed, although in all cases, the classes completed the test in 30–40 min. The students recorded their responses on the question paper. The students' responses were recorded and the percentage of students who responded to each alternative was calculated. The results of each class and school and the whole group were compared to determine whether the results were consistent or whether there were anomalies in the data.

The combinations of responses to the first and second tiers provide information about the students' reasoning in each item. Tan and Treagust (1999) suggested that when interpreting such data for the purpose of gaining an understanding of students' misconceptions, it is reasonable to consider those response combinations that exist for at least 10 % of students. The incorrect responses for which the percentage of students was greater than 10 % were further investigated to determine the types of reasoning used by the students. The class teachers and a sample of students were interviewed following administration of the test to identify any areas requiring attention, such as the wording of items or instructions. In addition, students were interviewed in focus

groups to identify any questions that they found confusing or unusual and to ensure that the responses were those that we had expected. The results of the pilot test have been reported elsewhere (see Hilton et al. 2012). The final version of the instrument is shown in Appendix 1.

As the final instrument was intended for use across Years 5 to 9, the items were designed to allow discrimination across year levels so it was expected that the students in the pilot study would experience difficulty in some situations (such as the inverse proportion question). Based on the data and feedback from the pilot, the original 12 items were retained for the final instrument. Some rewording was applied to make the meaning of the questions or responses clearer and three items were revised, however the majority of second-tier responses were retained.

The main study

The main study is being conducted over 3 years across 28 schools in six school clusters in South Australia and Queensland. Each cluster consists of two to five primary schools with one or two local secondary schools. In total, the study involves about 120 teachers and administrators and around 2100 students from Years 5 to 9 (10–14 years old). The consent of all participant students and their parents, as well as participating teachers, was sought prior to commencement of the study. The Queensland teachers have attended a series of eight half-day professional learning workshops and the South Australian teachers have attended a series of four full-day workshops. These were conducted over 2 years. At the beginning of the study, the teachers were asked to administer the final version of the diagnostic instrument to their classes. In the case of the secondary school teachers, the instrument may have been administered to more than one class, depending on the teacher's class allocation. The instrument was administered in class time and took each class 30–40 min to complete. The teachers were asked to read the items to their class if they considered it necessary, however, no elaboration on any item was given.

The purpose of the instrument was two-fold. First, it provided information to each teacher about the students in their own classes and the commonly used incorrect reasoning in situations of proportion and non-proportion. Second, it provided information for the research team about the problem types and reasoning employed by the students. This was used to inform the design of the professional learning workshops for the teachers.

Results

The data for the 140 students who participated in the pilot study showed that the majority of students were able to answer the non-proportional constant problem (Item 7 on the final instrument) correctly, however, for the majority of items, a high percentage of students chose an incorrect combination. These findings are consistent with those of the main study. The incorrect responses of students in the pilot study have been included in Table 4. The correct response data for the main study, arranged by year level, are shown in Table 2. These data reveal that while the number of students selecting the correct response generally increases with age, regardless of year level, all students found the same items similarly challenging.

The aggregated data for the full sample of 2,127 students in Table 3 indicate the number of students who selected each response combination.

Following recommendations suggested by Treagust and his colleagues (e.g., Peterson et al. 1989; Tan and Treagust 1999), who have used two-tier instruments extensively in science education research for many years, the combinations of incorrect responses selected by at least 10 % of the students were used to identify the common reasoning used by the students in each situation. The strategies used are shown in Table 4. This table illustrates the interpretation of the different response combinations. The data are presented in year levels to show trends in student reasoning. The results for the pilot test have been included where the items in both tests were closely aligned. The data show that the results for the students in the pilot study are generally consistent with those of the larger sample of students in the main study.

Interpreting the results of the individual items

The data shown in Table 2 indicate that, in general, regardless of year level, many students experienced difficulties with the same items. For example, fewer than 30 % of the students in any year level selected the correct response combination for eight of the items (2, 3, 5, 6, 9, 10, 11, 12). In the case of Item 12, less than 10 % of students in any year level answered correctly (two-dimensional scale). These findings align with those in the literature that proportional reasoning is difficult for many students and that it develops slowly. Indeed, for some people, even as adults, proportional reasoning is not easy. The item on which most students in each year level selected the correct combination was Item 7 (non-proportional constant situation). Even for this item, the number of students in Years 5 and 6 who answered correctly was only about 50 %. The results for this item improved greatly in Years 7 to 9, suggesting that students in these year levels were more adept at recognising a non-proportional constant situation. The results for the other non-proportional situation (Item 1) show that the students across all year levels experienced more difficulty in recognising a non-proportional additive situation than a constant situation. This reflects the results of a study by Van Dooren et al. (2005), which revealed similar patterns, both in terms of the trend across year levels and the percentages of students who responded correctly. Their study showed that additive strategies are typical of younger students and that for constant situations, the percentage of students who use additive strategies decreases with age. Further, the study also showed that the

Table 2 Percentage of students in each year level who selected the correct response combination

Year level	n	Item number											
		1	2	3	4	5	6	7	8	9	10	11	12
5	502	33.7	11.5	8.3	22.8	9.5	7.5	44.8	22.2	9.3	11.1	13.1	3.4
6	598	37.3	13.0	11.0	26.6	11.7	14.2	56.4	31.2	13.9	14.3	15.2	3.0
7	532	36.8	17.2	19.8	25.8	9.6	21.4	71.2	41.6	23.4	19.1	22.8	5.1
8	283	33.6	23.0	25.4	33.6	15.9	28.6	73.2	49.1	33.1	17.7	27.4	6.8
9	212	40.4	23.0	24.5	35.4	21.2	24.1	71.4	44.6	38.7	20.8	27.4	4.3

numbers of Year 5 to 8 students who correctly answered the additive situation were fairly similar, with a slight increase across the upper grades. Again, our results reflect these earlier findings.

It is perhaps unsurprising that for all items, as the year level increased, there was a general positive trend in the number of students who selected the correct response. The items that showed the greatest increase in correct response across the year levels were Items 2, 3, 5, 6, and 8–11, for which the percentage of students at least doubled from Year 5 to Year 9. These trends may be attributed to cognitive development of the students with age, as well as experience with and exposure to particular concepts through the curriculum.

The results in Table 3 are useful for identifying the commonly selected response combinations, which is an advantage of using such an instrument—investigating the response patterns allows identification of common errors within a group of students. It is interesting to note that only on four of the items was the percentage of correct responses higher than the percentages of students who chose other combinations. These items were the two non-proportional situations (Items 1 and 7), the item involving re-representation of a verbal distance–time situation graphically (Item 4), and the item about speed, a familiar rate relationship (Item 8). Some response combinations were unexpected; for example, a small percentage of students some-times selected a correct tier-1 response together with incorrect reasoning or, con-versely, an incorrect tier-1 response with correct reasoning. These percentages were generally less than 10 %.

Closer examination of the most common incorrect response combinations of those items to which less than 30 % of students responded correctly (i.e., Items 2, 3, 5, 6, 9–12) reveals clear patterns of erroneous reasoning that reflect those reported elsewhere in the literature. Each of these items is now considered in turn.

Table 3 Percentage of years 5 to 9 students (n=2,127) selecting each response combination

Item	Response							
	TA	FA	TB	FB	TC	FC	TD	FD
1	3.5	2.3	7.8	36.3[a]	30.7	5.4	10.8	3.3
2	1.9	7.8	7.7	30.9	16.4[a]	11.1	7.0	17.1
3	52.5	4.5	4.6	14.7	3.6	3.1	1.0	16.0[a]
4	15.1	11.4	32.3[a]	10.1	7.7	16.2	3.8	3.4
5	5.4	48.0	4.6	5.8	7.5	10.3	12.3[a]	6.1
6	18.4	4.3	6.6	46.8	17.5[a]	3.3	1.1	2.2
7	19.4	1.9	2.4	4.8	1.9	62.4[a]	4.4	2.8
8	14.9	3.8	2.5	35.8[a]	1.8	4.6	4.5	32.1
9	15.4	18.5	2.2	22.8	20.5[a]	10.7	5.1	4.9
10	7.5	16.2[a]	4.8	9.7	3.7	12.8	30.6	14.7
11	4.7	25.7	2.6	17.6	19.8[a]	3.2	2.9	23.6
12	8.2	4.3[a]	64.5	1.7	10.9	1.2	7.9	1.2

[a] Indicates the correct response combination for the item. TA, FA, etc. denote first- and second-tier response combinations

Table 4 Percentage of students using qualitative reasoning or erroneous strategies

Reasoning used	Choice	Pilot study	Main study by year level				
			5	6	7	8	9
Some qualitative understanding of relationship							
Understands need to compare part with whole	Item 5 (FC)	13.9	10.1	9.9	11.5	9.9	9.0
Qualitative understanding of speed-time relationship	Item 8 (FD)	33.6	35.1	27.6	33.0	33.1	31.5
Qualitative understanding of inverse relationship	Item 10 (D)	48.6	50.1	46.7	41.8	37.3	39.6
Some evidence of qualitative understanding of linear scale	Item 3 (FB)	10.7	9.4	15.0	17.6	17.0	15.1
Making absolute comparison instead of relative comparison							
Considers one value without considering the total	Item 5 (FA)	47.9	49.3	47.8	50.0	46.3	39.2
Considers one value without considering a related value	Item 6 (TA)	15.7	23.7	17.4	16.7	14.1	17.0
	Item 11 (FA, FB)	35.0	53.6	46.4	38.6	32.4	31.6
Reading scale as absolute, not considering value of units	Item 3 (TA)	58.6	56.6	53.9	50.3	48.4	44.8
Using multiplicative strategies inappropriately							
Applies a multiplicative strategy to a constant situation	Item 7 (TA)	13.6	24.1	19.7	16.7	15.4	16.0
Applies a multiplicative strategy to an additive situation	Item 1 (TC, TD)	21.4	27.5	40.1	43.9	48.4	41.8
Treats an inverse situation as if it involved direct proportion	Item 8 (TA)	12.1	20.5	18.9	11.2	5.7	10.3
Incorrect reasoning but recognises inverse situation	Item 10 (FC)	14.3	7.6	9.7	12.9	24.8	14.2
No recognition of impact of changing two dimensions on area	Item 12 (TB)	78.6	60.2	60.2	67.7	73.2	60.6
Using additive strategies in proportional situation							
Increases two quantities by the same amount (direct)	Item 2 (FB, FD)	56.5	43.0	47.1	53.0	45.0	44.6
Increases two quantities by the same amount (inverse)	Item 10 (FB)	10.0	9.5	9.1	9.6	9.2	10.4
Erroneous calculation/no calculation							
Inaccurately uses multiplicative reasoning	Item 2 (FC)	10.7	11.0	10.5	10.3	12.4	11.7
	Item 9 (FA, FB)		47.6	41.9	38.6	36.7	32.1
Not calculating a value/estimating	Item 6 (FB)	42.9	45.2	49.2	47.1	44.5	41.0
Faulty understanding of visual representations							
Interprets graph accurately but chooses irrelevant response	Item 4 (A)	47.2	38.5	37.3	34.7	32.2	37.0
Misinterprets a distance-time graph	Item 4 (FC)	10.0	13.0	12.4	16.7	15.5	9.0
Misinterprets relative amounts in pictorial representation	Item 11 (FD)	29.3	15.7	23.1	25.6	31.0	25.5

Item 2 required multiplicative thinking to correctly upscale two recipe components in proportion. The students needed to recognise that for a recipe with four cups of sugar and ten cups of flour, an increase to six cups of sugar requires a 50 % increase from 10 to 15 cups of flour. The percentage of correct responses was 16.4 %. The most common incorrect response (FB) showed that 30.9 % of students believed that since the cups of sugar increased by two, so too should the number of cups of flour be increased by 2. In addition to these students, a further 17.1 % of students selected FD. In this case, students focused on the absolute difference between cups of sugar and flour (you always need six more cups of flour). These response combinations indicate additive thinking and show that many students were not able to recognise the situation as one of proportion. The data in Table 3 show that this accounts for 43–53 % of students in the main study.

Item 3 showed a number line with a missing value. The correct answer, selected by 16 % of students, required recognition that the scale was 2 units for each segment, and that the unknown number was 8 more than the starting number. The most common absolute reasoning involved equating each segment with an increase of 1 unit, thereby valuing the unknown at 37. Students using this reasoning used absolute thinking rather than relative thinking. This response was selected by more than half of the students and, as shown in Table 4, was selected by only slightly less than 50 % of students in Years 8 and 9.

The correct response to Item 5 again required relative thinking and an understanding of the need to compare the part with the whole. This was selected by 12.3 % of the students. Contrasting with this, the most common response, selected by 48 % of students, was FD, indicating that students considered only the absolute numbers in the two classes who voted for each activity. They did not consider the fact that one class had far fewer students in total than did the other. There was a small decrease from Year 8 to 9 in the percentage of students who selected this response.

A similar pattern emerged in the responses to Item 6. This item required students to compare the cost of washing powders A and B relative to the number of washing loads to determine that Powder A was more economical since the cost per load was less than that of Powder B. The correct response was selected by 17.5 % of students. The most common incorrect response, chosen by 46.8 % of students, was FB. This response stated that Powder B costs a little more but provides ten more loads of washing. It again suggests absolute thinking since the only comparison involves the loads of washing. This response was also found to be the most common on the KIIP test used by Dole et al. (2007).

Item 9 is somewhat problematic. The correct response clearly indicates that students are able to recognise a missing-value situation and use multiplicative thinking to make a correct calculation. This was achieved by 20.5 % of the students. However, while the incorrect response combinations indicate that students were unable to do this, they reveal little about the reasons for this. Almost 23 % of students selected FB, which suggests that they made an erroneous calculation and arrived at an answer whose value was less than that of the correct answer. Perhaps this might be the case had these students employed additive thinking, although there is no evidence for this without access to their working. For these reasons, we feel that this item requires amendment or omission from the instrument.

Item 10 focused on an unfamiliar situation involving inverse proportion. The responses to this item are interesting because they suggest that the majority of students

recognised the situation as an inverse proportion. In addition to the 16.2 % who selected the correct response combination, 45.3 % of students selected either FD or TD together, which indicates a qualitative level of understanding of the relationships between time and people (reason D states, "There is less time so more people are needed"). Of the remaining students, those who selected reasons B or C have used additive reasoning. Reason B still indicates some understanding that as time available decreases, the number of people increases, however, reason C suggests that students have not recognised this inverse relationship. The number of students who answered this item correctly increased with the age of the students, which may indicate an effect related to curriculum.

The incorrect responses to Item 11 fall into two large groups. A total of 43.3 % of students selected FA or FB. These students have used absolute thinking, comparing only the amount of sugar (FA) or the amount of water (FB), but not making any relative comparison between the two. A further 23.6 % of the students selected FD. In this case, the students chose the statement that all three cups were of equal sweetness. These students may have incorrectly interpreted the visual representation or perhaps having some misconceptions around equivalent fractions, since Cup A contains 3/3 cup of water and 3 sugar cubes; Cup B contains 1/2 cup of water and 2 sugar cups; and Cup C contains 1/3 cup water and 1 sugar cube.

Item 12 revealed a situation in which most students made an erroneous assumption. The majority of the students (64.5 %) selected TB, which indicates that they have only considered the change in one dimension, rather than recognised that a doubling of both dimensions will result in an area four times the original size.

Overall trends

The response patterns for Items 3, 5, 8, and 10 suggest that the students had some level of qualitative understanding of the situation or relationship. This is shown in the first section of Table 4. In these cases, the students' responses showed some understanding of the co-variation involved in the situation, however, they did not select the response that involved the quantitative answer. According to Cramer and Post (1993), qualitative thinking is an important part of proportional reasoning. They argued that including situations involving qualitative thinking in teaching is important for developing students' problem solving in proportional situations. The students' qualitative approaches revealed by the instrument suggest that more targeted work with such students might assist them to develop a deeper understanding.

An examination of the other responses of students revealed clear patterns of erroneous reasoning, which align closely with those described in the literature. In line with findings of several researchers (e.g., Cramer et al. 1993; Van Dooren et al. 2005, 2010), the results show that the students had problems discriminating non-proportional from proportional situations. In both non-proportional situations (Items 1 and 7), the majority of students who chose incorrectly used multiplicative thinking, which suggests that they identified the situations as proportional.

In contrast, and as reported frequently in the research literature, in some proportional situations (e.g., Items 2 and 10), the majority of incorrect responses indicated the use of an additive strategy when multiplicative thinking was required. In these cases, the students have not identified the situations as involving multiplicative relationships. Item 2 involved direct proportion, and Item 10 involved inverse

proportion. In other proportional situations, many students used absolute comparison in situations requiring relative comparison, either comparison of part-to-part or part-to-whole (Items 3, 5, 6, 11). The scale-related questions required multiplicative strategies (two-dimensional) or relative thinking (one-dimensional), and these were also problematic for many students.

Finally, it appears that students are comfortable with some mathematical representations (e.g., graphs) but that they do not always interpret visual representations correctly or apply relative thinking in such situations accurately.

Discussion

This paper describes an instrument that can be used across a range of year levels to diagnose students' proportional reasoning, application of multiplicative and additive thinking, or their use of relative and absolute thinking in a range of problem types. The results show that the instrument does provide insights into reasoning used by students in both proportional and non-proportional situations. The findings are consistent with the research literature and the instrument was successful in identifying the same trends as reported from other studies.

The teachers and the researchers have used the data from the instrument in a number of ways during the study. Research suggests that teachers may need support in developing strategies and activities to assist their students to develop sound skills and conceptual understanding associated with proportional reasoning (Sowder et al. 1998). The instrument was a useful tool for designing professional development for the teachers involved in our study, whose students completed the instrument. The results of the instrument allowed us to respond directly to the students' understanding of proportional and non-proportional situations by designing classroom activities for use in the professional development workshops that could be adapted by the teachers to target particular needs of their students.

The instrument also provided the teachers with data about their own students. The research team coded the responses and the results for each class were collated and presented to the teachers. Each teacher received the data for their class or classes in a form similar to Table 2 with a second table that showed the response combinations of individual students. Discussions were held with the teachers to help them interpret the data and to identify the reasoning commonly used by their students. Because the instrument was designed for use with a range of year levels (from Year 5 to Year 9), the teachers were encouraged to consider those items or situations most appropriate to the ages and curriculum requirements of their students. For example, it may not be appropriate to expect Year 5 students to have a quantitative understanding of the inverse situation presented in Item 10. Similarly, those responses indicating a qualitative level of understanding were identified as a starting point for target teaching to help students develop an understanding of the situations at a quantitative level. The teachers used the findings to plan and select activities that might strengthen their students' understanding or address areas of weakness identified by the items.

The instrument has also proved valuable as a diagnostic tool for the teachers themselves. According to Sowder et al. (1998), teachers of middle years students are not always well equipped to help students to develop proportional reasoning because many teachers have themselves been inadequately instructed and

consequently have limited proportional reasoning strategies. Through administering the test to their students and analysing the response combinations in professional development workshops, the teachers became cognisant of their own understanding in addition to the areas of difficulty for their students. Sowder et al. (1998) made a number of recommendations associated with developing teachers' understanding of multiplicative structures. These included providing teachers with opportunities to explicitly use proportional reasoning in a range of situations and contexts. In order to assist the teachers to develop their understanding of the concepts and strategies involved in situations of proportion and multiplicative thinking, hands-on workshop activities were designed to engage the teachers in problem solving and reflection on their own strategies. The activities were designed so that the teachers could adapt them for use in their classrooms to help them to target a number of problem types and proportional reasoning strategies that the data from the instrument identified as problematic. These included:

- distinguishing non-proportional and proportional situations
- identifying when it is appropriate to use additive and multiplicative strategies
- distinguishing relative and absolute situations
- identifying situations in which relative thinking is required and applying relative thinking to make part-whole comparisons
- working with scale in one, two, and three dimensions—situations that involve relative thinking and investigating the influence of changing one or more dimensions on properties such as area or volume
- strategies for solving proportional problems (e.g., ratio tables; physical tools such as balance beams, counters, blocks)
- development of basic multiplicative skills and knowledge, such as fractional thinking, place value, percentages
- activities associated with common rates, e.g., speed and density, and associated sets, e.g., surface area-volume relationships
- using multiple representations to express mathematical relationships
- using digital technologies and images to foreground proportional situations and elicit discussion with students

The activities were presented using a variety of contexts and subject areas. For example, during 2012, contexts included the sinking of the Titanic, the Transit of Venus, and the Olympics. Those subject areas with which the activities were aligned included history, geography, science, mathematics, and physical education. In some cases, activities were designed at the school or cluster level to align with programs already running in those schools. For example, in one cluster, many schools have a kitchen garden program. The teachers approached the research team to design proportional reasoning teaching resources for use during lessons held in the kitchen and garden (see Hilton et al. 2013).

Conclusion

The proportional reasoning two-tier instrument described in this paper is a new approach to assessing students' proportional reasoning. The data generated from this two-tier instrument allow students' common errors or partial understanding in proportional and

non-proportional situations to be identified, which informs researchers and teachers about the areas that teachers might target to enhance students' proportional reasoning. For example, those items that indicate students' use of absolute comparisons when they should make relative comparisons inform teachers that this is an area requiring further targeted work with their students. Such data may also be useful for teacher educators to inform the design of teacher professional development to promote teachers' understanding of the different types of proportional reasoning and classroom strategies to address their students' needs.

The instrument described in this paper has been used by the teachers participating in the study as a pre-test to identify students' understanding and strategies and as a post-test to track students' progress and to identify changes in their proportional reasoning. This has informed the teachers' planning and practice. In light of the importance of proportional reasoning development and the difficulties students encounter when reasoning proportionally, there is a clear need to identify those problem types and reasoning skills with which students require support. The various applications of this instrument have the potential to benefit teachers, curriculum planners, researchers, teacher educators, and, most importantly, students.

Appendix 1 The proportional reasoning diagnostic instrument

Student Instructions:

1. Each question has two parts. In the first part, circle either *True* or *False*.
2. In the second part, circle the letter that matches the best reason (*A, B, C* or *D*).
3. Please use only pencil. If you make a mistake, completely erase it and try again.

Note: Font size and spacing have been adjusted for presentation in this article.

Question 1 Running laps

Sue and Jack were running equally fast around a track. Jack started first. When Jack had run 4 laps, Sue had run 2 laps.
When Sue had completed 6 laps, Jack had run 12 laps.
True or False

Because (choose the best reason)

A. The further they run, the further Jack will get ahead of Sue.
B. Jack is always 2 laps ahead of Sue.
C. Jack completes double the laps of Sue.
D. Sue has run 3 lots of 2 laps to make a total of 6 laps, so Jack must have run 3 lots of 4 laps to make a total of 12 laps.

Question 2 A sticky mess

To make a sticky mess, you need to mix 4 cups of sugar and 10 cups of flour. If you wanted to make a larger amount of this recipe with 6 cups of sugar, you would need 15 cups of flour.

True or False

Because (choose the best reason)

A. You don't need to change the amount of flour.
B. You have 2 more cups of sugar so you need 2 more cups of flour.
C. You increased the sugar by half so you need to increase the flour by half.
D. You always need six more cups of flour than sugar.

Question 3 A number line

33   X   53
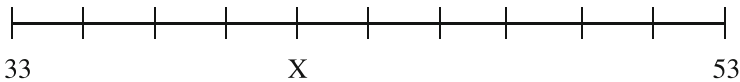
On this number line, X represents 37.
True or False

Because (choose the best reason)

A. X is 4 cm along the line and 33+4=37.
B. X needs to be closer to 33 than 53.
C. X is nearly half-way so it looks OK.
D. X is 8 more than 33.

Question 4 Riding home with Jane

Jane rode her bicycle home. She rode at a steady speed for a short time and then she had a rest. After her rest, she rode at double her original speed. She drew a graph to represent her journey. Jane's graph below is correct:
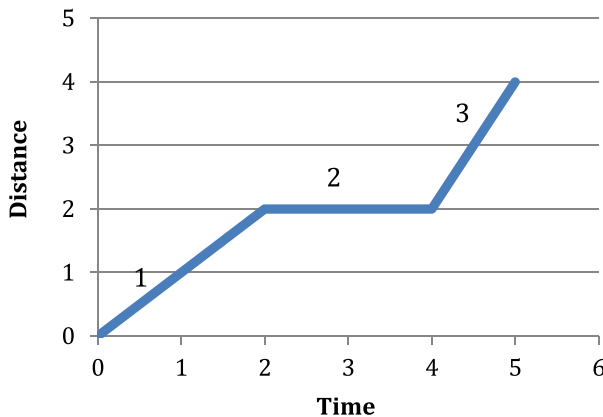
True or False

Because (choose the best reason)

A. The distance covered in Part 3 is greater than the distance in Part 1.
B. Part 3 is twice as steep as Part 1.

C.   Part 3 needs to be twice as long.
D.   The times for parts 1 and 3 are the same.

Question 5 End-of-term activities

This table shows the end-of-term activities voted by Year 5 and Year 6 students.

Year Level	Students who chose the beach	Students who chose the movies	Total Students
Year 5	8	14	22
Year 6	7	6	13

Going to the beach is a relatively more popular choice with the Year 6 students than the Year 5 students.
True or False

Because (choose the best reason)

A.   More students in Year 5 chose the beach.
B.   Only 6 students in Year 6 chose not to go to the beach.
C.   Fewer students in Year 6 chose the beach but there are fewer students in the class.
D.   More than half of the Year 6 students chose the beach and less than half of the Year 5 students chose the beach.

Question 6 Washing days

Washing Powder A:
- $4.00
- 1 kg
- 20 loads of washing

Washing Powder B:
- $6.50
- 1.5 kg
- 30 loads of washing

Washing powder A is the best value.
True or False

Because (choose the best reason)

A.   Washing powder A costs the least.
B.   Washing powder B costs a little bit more but you get 10 more loads of washing.
C.   The cost per load of washing is less.
D.   Both washing powders are the same value.

Question 7 Sing Song

A group of 5 children takes 4 min to sing a song. A group of 10 children would take 8 min to sing the same song.

True or False

Because (choose the best reason)

A.   Doubling the number of children would double the time to sing the song.
B.   Double the children should halve the time to sing the song.
C.   The number of children does not affect the time to sing the song.
D.   Adding more children increases the time taken to sing the song.

Question 8 Speedy George

George runs 100 m in 20 s. If he runs the same distance at twice the speed, he will take twice as long.

True or False

Because (choose the best reason)

A.   Doubling the speed doubles the time.
B.   Doubling the speed halves the time.
C.   The distance doesn't change.
D.   Running faster will take less time.

Question 9 Funky music

Sandra decided to save to buy an iPod costing $84. To help her buy the iPod, Sandra's mother agreed to give her $5 for every $2 that Sandra saved.

Sandra has saved $24, so she has enough for the iPod.

True or False

Because (choose the best reason)

A.   Now Sandra and her mother have more than they need for the iPod.
B.   Now Sandra and her mother still don't have enough for the iPod.
C.   Sandra's mother will give Sandra $60.
D.   As long as Sandra's mother pays more, it's OK.

Question 10 Fence Painting

Six people can paint my fence in 3 days. If all people paint at the same rate, it would take 12 people to paint the fence in 2 days.
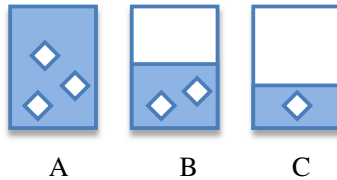
True or False

Because (choose the best reason)

A.   The number of people multiplied by the number of days must stay the same, so you need 9 people.
B.   If you decrease the time by 1 day, you must increase the people by 1, so you need 7 people.
C.   If you decrease the time by 1 day, you must decrease the people by 2, so you need 4 people.

D.   There is less time so more people are needed.

Question 11 Three cups

Three cups have different amounts of water and sugar. Cup A is full of water with 3 lumps of sugar. Cup B is half full of water with 2 lumps of sugar. Cup C is one third full of water with 1 lump of sugar.
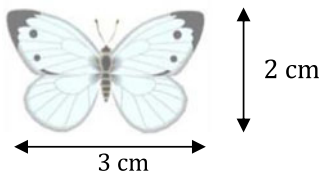

A   B   C

When the lumps of sugar have been stirred in, Cup B will be the sweetest
True or False

Because (choose the best reason)
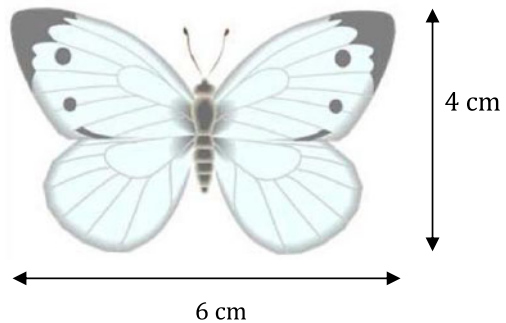
A.   Cup A is the sweetest because it has the most sugar.
B.   Cup C is the sweetest because it has the least water.
C.   A full glass of B would need 4 lumps of sugar.
D.   They are all the same sweetness.

Question 12 Drawing insects

Insect A: **Insect B:**

2 cm

3 cm

4 cm

6 cm

Bill has drawn two diagrams. The area of insect B is twice that of insect A.
True or False

Because (choose the best reason)

A.   The area of Insect B is 4 times greater.
B.   Insect A is half the width of Insect B.
C.   Insect B is twice as long as insect A.
D.   Bill has only doubled one dimension.

References

Ahl, V. A., Moore, C. F., & Dixon, J. A. (1992). Development of intuitive and numerical proportional reasoning. *Cognitive Development, 7*, 81–108.

Akatugba, A. H., & Wallace, J. (2009). An integrative perspective on students' proportional reasoning in high school physics in a West African context. *International Journal of Science Education, 31*(11), 1473–1493.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*(5), 1478–1490.

Bright, G. W., Joyner, J. M., & Wallis, C. (2003). Assessing proportional thinking. *Mathematics Teaching in the Middle School, 9*(3), 166–172.

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2008). An evaluation of a teaching intervention to promote students' ability to use multiple levels of representation when describing and explaining chemical reactions. *Research in Science Education, 38*, 237–248.

Cramer, K., & Post, T. (1993). Making connections: A case for proportionality. *Arithmetic Teacher, 60*(6), 342–346.

Cramer, K., Post, T., & Currier, S. (1993). Learning and teaching ratio and proportion: Research implications. In D. Owens (Ed.), *Research ideas for the classroom* (pp. 159–178). NY: Macmillan Publishing Company.

Dole, S., Wright, T., Clarke, D., & Hilton, G. (2007). Making connections science and mathematics: The MCSAM Project. In U. Cheah, Y. Wahyudi, R. Devadason, K. Ng, J. Chavez, & D. Mangao (Eds.), *Redefining learning culture for sustainability* (pp. 184–194). Second International Conference on Science and Mathematics Education, Penang, Malaysia, 13–16 November 2007.

Haslam, F., & Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of Biological Education, 21*, 203–211.

Hilton, A., Hilton, G., Dole, S., Goos, M., & O'Brien, M. (2012). Evaluating middle years students' proportional reasoning. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics education: Expanding horizons* (pp. 300–307). Proceedings of the Thirty-fifth Annual Conference of the Mathematics Education Research Group of Australasia. (MERGA-35). Singapore: MERGA.

Hilton, A., Hilton, G., Dole, S., Goos, M., & O'Brien, M. (2013). Kitchen gardens: Contexts for developing proportional reasoning. *Australian Primary School Mathematics Classroom, 18*(2), 21–26.

Howe, C., Nunes, T., & Bryant, P. (2010). Rational number and proportional reasoning: Using intensive quantities to promote achievement in mathematics and science. *International Journal of Science and Mathematics Education, 9*, 391–417.

Lamon, S. J. (1993). Ratio and proportion: Connecting content and children's thinking. *Journal for Research in Mathematics Education, 24*(1), 41–61.

Lamon, S. J. (2005). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers*. Mahwah, NJ: Lawrence Erlbaum.

Langrell, C. W., & Swafford, J. (2000). Three balloons for two dollars: Developing proportional reasoning. *Mathematics Teaching in the Middle School, 6*(4), 254–261.

Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 93–118). Reston, VA: Lawrence Erlbaum & National Council of Teachers of Mathematics.

Misailidou, C., & Williams, J. (2003). Diagnostic assessment of children's proportional reasoning. *Journal of Mathematical Behaviour, 22*, 335–368.

Modestou, M., & Gagatsis, A. (2007). Students' improper proportional reasoning: A result of the epistemological obstacle of "linearity". *Educational Psychology, 27*(1), 75–92.

Modestou, C., & Gagatsis, A. (2010). Cognitive and metacognitive aspects of proportional reasoning. *Mathematical Thinking and Learning, 12*, 36–53.

Nabors, W. (2003). From fractions to proportional reasoning: A cognitive schemes of operation approach. *Journal of Mathematical Behaviour, 22*(2003), 133–179.

O'Keefe, L., & O'Donoghue, J. (2011). The use of evidence based research on mathematics textbooks to increase student conceptual understanding. *International Journal for Cross-Disciplinary Subjects in Education, 2*(1), 304–311.

Özmen, H. (2008). The influence of computer-assisted instruction on students' conceptual understanding of chemical bonding and attitude toward chemistry: A case study for Turkey. *Computers and Education, 51*(1), 423–438.

Park, J. S., Park, J. H., & Kwon, O. N. (2010). Characterizing the proportional reasoning of middle school students. *Seoul National University Journal of Education Research, 19*(5), 119–144.

Peterson, R. F., Treagust, D. F., & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate Grade-11 and -12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching, 26*(4), 301–314.

Sowder, J., Armstrong, B., Lamon, S., Simon, M., Sowder, L., & Thompson, A. (1998). Educating teachers to teach multiplicative structures in the middle grades. *Journal of Mathematics Teacher Education, 1*, 127–155.

Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education, 23*, 285–292.

Tan, K.-C. D., & Treagust, D. F. (1999). Evaluating students' understanding of chemical bonding. *School Science Review, 81*, 75–83.

Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics, 16*, 181–204.

Treagust, D. F. (1995). Diagnostic assessment of students' science knowledge. In S. M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 327–346). Mahwah, NJ: Lawrence Erlbaum.

Treagust, D. F. (2006). *Diagnostic assessment in science as a means to improving teaching, learning, and retention*. Paper presented at the UniServe Science Symposium Proceedings Assessment in Science Teaching and Learning, Sydney, NSW, Australia.

Tüysüz, C. (2009). Development of two-tier diagnostic instrument and assess students' understanding in chemistry. *Scientific Research and Essay, 4*(6), 626–631.

Van De Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics: Teaching developmentally*. Boston, MA: Allyn & Bacon.

Van Dooren, W., De Bock, D., & Verschaffel, L. (2010). From addition to multiplication … and back: The development of students' additive and multiplicative reasoning skills. *Cognition and Instruction, 28*(3), 360–381.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralisation. *Cognition and Instruction, 23*(1), 57–86.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.