UNIVERSITY OF CAMBRIDGE

Surveying mathematics teachers' knowledge of formative assessment: a study of teachers in the Federal District of Brazil

Melise Maria Vallim Reis Camargo

Girton College

This dissertation is submitted for the degree of Doctor of Philosophy.

Supervisor: Prof Kenneth Ruthven
Advisor: Dr Andreas Stylianides
Submission: August, 2018

To Rafael

My partner, my best friend, my soulmate

ABSTRACT

The research presented in this dissertation involves a quantitative study of mathematics teachers' knowledge of formative assessment. Formative assessment is understood as a process in which both teachers and students actively become the agents of the process, responsible for their own knowledge and practice. In this process, *formally* gathered evidence is used to formulate feedback and inform decisions; and *informal* evidence (e.g. observation, conversations) is used to generate teacher and peer-feedback to improve learning (Hargreaves, 2005; Wiliam & Leahy, 2015).

The focus of the study is on mathematics teachers in state secondary schools in the Brazilian Federal District. The research design employs survey methodology with a structured e-questionnaire. The questionnaire was designed based on six domains of knowledge extracted from existing research literature. The several piloting phases, and a field test conducted with a larger sample, demonstrate the validity and reliability of the instrument and provide information about mathematics teachers' knowledge of formative assessment.

The evidence shows that teachers in the Federal District did relatively better in terms of interpreting evidence of students' learning and helping students to use assessment information. On the other hand, they had a relatively lower performance in terms of choosing/developing assessment methods (e.g. classroom activities, discussions) to elicit evidence of students' learning. The overall performance of teachers in the Federal District was lower than that of teachers sampled from the other states of Brazil.

The original contribution of this research is *methodological* in the development, piloting and application of a new instrument to assess mathematics teachers' knowledge of formative assessment; and *to knowledge* in providing information and a unique insight into Brazilian mathematics teachers' knowledge of formative assessment. There are important implications for policy and practice, focussing on teachers' professional development with regards to formative assessment and clarifying Brazilian teachers' roles as assessors.

RESUMO

A pesquisa aqui apresentada envolve um estudo quantitativo sobre o conhecimento dos professores de matemática sobre avaliação formativa. Esta é entendida como um processo em que professores e alunos são agentes ativos responsáveis por seus próprios conhecimentos e práticas. Neste processo, evidências formalmente coletadas são usadas para formular feedback e informar decisões; e evidências informais (como observação, conversas) são usadas para professores e colegas fornecerem feedback para melhorar a aprendizagem (Hargreaves, 2005; Wiliam & Leahy, 2015).

O foco deste estudo são professores de matemática de escolas públicas secundárias do Distrito Federal brasileiro. A pesquisa utilizou *survey* como metodologia e um questionário eletrônico estruturado como método de coleta de dados. O questionário foi elaborado com base em seis conhecimentos extraídos da literatura existente. Várias fases de teste realizadas com professores do Distrito Federal e com uma amostra maior dos outros estados do país demonstram a validade e a confiabilidade do instrumento.

Os resultados mostram que professores do Distrito Federal tiveram uma performance relativamente melhor em relação à interpretação do aprendizado dos alunos e em ajudar os estudantes a utilizar as informações fornecidas pelas avaliações. Por outro lado, tiveram um desempenho relativamente menor em relação à escolher ou desenvolver métodos de avaliação (como atividades de sala de aula, discussões) para coletar evidências de aprendizado dos alunos. O desempenho geral dos professores no Distrito Federal foi menor do que o da amostra de professores dos outros estados do Brasil.

A contribuição original desta pesquisa foi *metodológica* no desenvolvimento, teste e aplicação de um novo instrumento para avaliar o conhecimento dos professores de matemática sobre avaliação formativa; e também em fornecer informações únicas sobre o conhecimento sobre avaliação formativa dos professores de matemática brasileiros. Implicações importantes para políticas públicas e práticas são fornecidas, com foco no desenvolvimento profissional dos professores no que se refere à avaliação formativa e à necessidade de esclarecimento do papel dos professores como avaliadores.

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed the prescribed limit of 80,000 words.

Melise Maria Vallim Reis Camargo
August, 2018

Acknowledgements

Once I read that doing a PhD is not just writing clever thoughts into chapters. It is a personal endeavour that demands us to rewrite our personhood too. In my case, this involved the generous support of many individuals and institutions.

I owe special thanks to my supervisor, Ken, whose patience, expertise and invaluable comments guided me through the process.

I am also grateful to my wonderful friends Leila, Aline, Nina, Hanan and Farida. Leila, for her hands-on help on feeding back, writing and re-writing some of the questions of the questionnaire with me; Aline, Nina, Hanan and Farida for simply being there – any time and any place I needed.

In addition, I thank all people who pointed me toward relevant texts, answered questions, read drafts, helped me with translations and statistics, offered advice, or gave me a place to stay while conducting research. This group includes: Caio, Karen, Ratna, Fiona, Ana Porto, Carine, Carmyra, Cília, Cleyton, Cristiano, Deire, Francisco, Marilene, Raimunda, Raquel Dörr, Raquel Santana, Rejane, Amaral, Ana Paula, Cacilda, Cibelle, Jorge, Manoel, Marcus, Marlene, Benigna, Cláudia, Enílvia, Erisevelton, Maria Suseley, Rose Meire, Sandra, Terezinha, Pam and Sue.

I am especially grateful to the eight teachers and the 19 researchers who contributed in the validation phases, the 232 teachers who participated in the field test and the 169 mathematics teachers from the Federal District who answered the final version of the questionnaire. Without their participation, this research would never happened.

I also would like to thank my Brazilian-Cambridge family for their solidarity throughout this project: Débora, Simon, Cris, Leo, Tati, Be, Emilene, Gabriel, Ana, Melissa and Alexandre.

In addition, I am grateful for the financial support of Capes and the Cambridge Trust, and my colleagues at Cambridge Assessment for their support and encouragement. A special thank you also goes to the staff of the Faculty of Education and Girton College.

My deepest thanks, however, I owe to two people:

My husband and best critic, Rafael, for his input about everything from statistics analysis to polishing the writing. While his copious feedback made the completion of this project possible, in a practical sense more generally his ability to never lose track of the big picture, and his unconditional love and support throughout the years, helped me keep going.

My dear friend Sarah who dedicated uncountable days (including weekends) to read and re-read my dissertation with me, helping me edit and proof-read the drafts. I would never be able to write my ideas in such high-quality English if it wasn't for her. She did it in such a caring, supportive and fun way that I will never be able to thank her enough.

Finally, I cannot resist thanking Gael and Ana Laura, for existing and being the light of my life; and my parents Vera and Iracino, my sister Milene and her husband Fabinho, who gave me my precious nephews Theodoro and Benício, for reminding me everyday of how lucky and blessed I am to be born in this loving and caring family.

Por último, eu não posso deixar de agradecer ao Gael e à Ana Laura por simplesmente existirem e serem a luz da minha vida; e aos meus pais Vera e Iracino, minha irmã Milene e meu cunhado Fabinho, que me deram meus preciosos sobrinhos Theodoro e Benício, por me lembrarem todos os dias o quão sortuda e abençoada eu sou por ter nascido nesta família tão amorosa e atenciosa.

CONTENTS

List of Tables

List of Figures

List of Abbreviations

df degrees of freedom 194, 195

AfL Assessment for Learning 38, 39

ANEB National Assessment of Basic Education 33

ANRESC National Assessment of Educational Achievement 34

CPD Continuing Professional Development 25, 83, 251, 255

CTT Classical Test Theory 245

DoK Domains of Knowledge 84–86, 91–93, 95, 96, 99, 103, 104, 111, 119, 122, 123, 140, 144, 145, 186, 193, 207–210, 213, 223, 238, 241, 243, 244, 249, 253, 256

EFA Exploratory Factor Analysis 142, 192, 193

ENEM National Middle Education Exam 35

FA formative assessment 21, 22, 24–26, 36–39, 41, 52, 58, 62, 64, 69–71, 73–77, 79, 80, 82, 84–86, 88–93, 95–99, 103, 107, 110, 112, 113, 119, 122, 144, 146, 161, 171, 186, 187, 193, 194, 196, 197, 204, 206–208, 219, 221–226, 230, 231, 234, 235, 237–240, 242–259

FD Federal District 22, 25–30, 32, 34, 37, 63, 64, 69, 72, 74, 81–83, 111–113, 124–126, 140, 157, 170, 172, 178, 187–194, 196–198, 204–206, 209, 211–216, 219, 221–223, 226–232, 235–248, 250, 252–255, 258, 259, 285, 395

GEA Guidelines of Educational Assessment 69–74, 81, 231

ICT Information and Communication Technology 68, 124

IDEB Index of Basic Education Development 34

INEP National Institute of Educational Studies and Research 33

IRT Item Response Theory 245

KAM Knowledge of designing and/or choosing assessment methods or classroom activities and discussions to collect evidence of students' learning 85, 88, 90, 149, 166, 170, 208, 210, 212, 230–232, 238, 253, 285, 287

KCL Knowledge of closing the feedback loop 86, 93, 108, 160, 176, 185, 208, 210, 226, 286, 287

KEF Knowledge of providing effective feedback 85, 87–91, 106, 121, 158, 174, 183, 208–210, 216, 232, 238, 286, 287

KHS Knowledge of helping students use assessment information 86, 155, 172, 208, 209, 213, 216, 226, 238, 241, 287

KIE Knowledge of interpreting evidence of students' learning 85, 87, 88, 105, 152, 163, 181, 208, 209, 213, 216, 224, 225, 230, 238, 254, 286, 287

KLI Knowledge of articulating and sharing clear learning intentions and success criteria 85, 87, 89–91, 153, 168, 179, 208, 210, 211, 228, 238, 254, 285–287

KMO Kaiser-Meyer-Olkin 142, 193

KMOFAP King's-Medway-Oxfordshire Formative Assessment Project 51

KOSAP King's Oxfordshire Summative Assessment Project 53

LDB-96 General Law of Education 1996 27, 30, 64

MaTFAKi Mathematics Teachers' Formative Assessment Knowledge instrument 77–82, 84–87, 92, 95–101, 103–105, 110, 112, 113, 116–118, 120–122, 124–126, 130, 132, 133, 139–141, 143, 144, 146, 148, 157, 171, 184, 186–190, 193, 197, 198, 202–204, 206–208, 221–224, 226, 227, 230, 231, 234, 238, 239, 241–250, 254, 256–258

OMC ordered multiple-choice 97, 105, 110, 112, 223, 224

PCA Principal Component Analysis 142

PCN National Curriculum Parameters 64, 67

SAEB Assessment System of Basic Education 33

SJT Situational Judgement Test 97–99, 143–145, 190

CHAPTER 1

INTRODUCTION

In 1976, British philosopher and author Celia Green wrote: “The way to do research is to attack the facts at the point of greatest astonishment”. As a secondary mathematics teacher in Brazil, my point of greatest astonishment was the impact of the ways my colleagues and I were accustomed to assessing our students and how this could affect their learning of mathematics.

This chapter outlines how this astonishment turned into research on secondary mathematics teachers’ knowledge of formative assessment (FA). The aim is to explain its importance and my motivation to conduct the research discussed in this dissertation.

1.1 Is she going to assess us like you do?

I have been teaching since 2006, when I completed my degree in mathematics. During this period, I not only taught in secondary schools, but I had the opportunity to work in three different environments which gave me multiple views of the landscape of mathematics education in Brazil.

In both state and private secondary schools, I experienced what it means to be a mathematics teacher and to deal with the daily routine of a classroom. As a college teacher and a Master’s student, I had the opportunity to see how undergraduate students behave in different situations involving problem solving and dealing with different points of view during different activities; and how they were influenced by the pedagogical orientation and assessment practices which their school teachers tended to follow. Finally, as a teacher trainer, I was in constant contact with various practitioners and their expectations and concerns about mathematics teaching and learning.

Through these different experiences, I developed a more extended view about assessment and how this is a subject of major concern to teachers. However, it was one particular episode in 2009 that drew my attention to the importance of assessment.

I was teaching in one of the poorest areas of the Brazilian Federal District (FD), when I was invited to be the vice-principal of my previous school. On the week of my leaving, I decided to talk to all my students, to explain why I would not be their mathematics teacher any longer.

When I mentioned that another teacher was already coming to replace me and that everything would be alright in the end, one student raised his hands and asked, "Is she going to assess us like you do?" This simple question made me raise a million others. Why he was asking me this? Of all the questions he could have asked, why was assessment the topic of his concern?

This personal experience made me think about my assessment practices and why they were so remarkable for him. I did not know the answer. If I was not able to answer the question about my own practices, what about other teachers? How were they assessing their students? Did they understand the principles of FA? Were they aware of their own practices? Did they know how to develop or select the appropriate method for assessing their students? Did they have the skills to interpret the results of their assessment and then adjust their instruction in the light of them? These are just some questions that arouse in my own mind when thinking about assessment, and which have been the subject of intense research in recent years due to its importance (Black & Wiliam, 1998; Kingston & Nash, 2011; Laveault, 2016; McMillan, 2013; Shepard, 2013; Stiggins, 2002; Veldhuis & van den Heuvel-Panhuizen, 2014; Xu & Brown, 2016).

1.2 The importance of assessment

Assessing students is an important role of classroom teachers as it contributes to every other teacher function (Brookhart, 1998, 1999). It has been estimated that teachers spend up to 50% of their time on assessment-related activities (Plake, 1993; Stiggins, 1999).

However, it is safe to say that until very recently, classroom assessment followed the same summative pattern: teachers would teach and then test what their students had grasped of that content. Their judgement would then be based only on the results of these tests. The next step would be to move on to the following content set out in the curriculum. More recently, however, changes in society have had a major impact on the expectations of students' learning, challenging this model.

Added to that, it has been proved that assessment has a considerable impact on students' learning (Black & Wiliam, 1998) and that there are significant relationships between students' performance and teachers' assessment practices (Rodriguez, 2004). This impact is not just on students' performance, but on students' emotions and their motivation as learners (Brookhart, 1997; Stiggins, 2002, 2007).

It is also the view of some that to help students develop 21st-century competencies such as critical thinking, problem solving, collaboration, communication, and self-regulated learning, teachers need to be equipped with the knowledge and skills to implement quality assessment (Koh et al., 2015; Shepard, 2013).

This can prove to be very problematic, as many studies have already shown that teachers' assessment skills are generally weak (Brookhart, 2001; Campbell, Murphy, & Holt, 2002; Campbell, 2013; Plake, 1993). Black and Wiliam (1998), for example, discussed the results of studies conducted by Crooks (1988) and Black (1993), which revealed that:

- existing teacher practices generally encouraged superficial and mechanical learning, focusing on memorising isolated details that students would forget quickly;
- teachers, in general, did not critically analyse their use of assessment tasks and procedures or discussed them with students, which indicated little reflection on what was being assessed;
- the attribution of marks was the primary purpose rather than the promotion of learning;
- there was a tendency to conduct norm-referenced rather than criterion-referenced assessment, which emphasised competition among students, leaving aside their individual development. This practice meant that feedback was used primarily to inform weaker students that they lacked skills; discouraging them from believing in their own ability to learn.

DeLuca and Klinger (2010) and Volante and Fazio (2007) have also shown that most teacher candidates reported a low level of assessment knowledge and skills, especially in

FA. This can undermine teacher confidence about assessment (Scott, Webber, Aitken, & Lupart, 2011), which in turn can result in assessing students improperly and inaccurately (Stiggins, 2001).

Based on this argument, many pieces of research have tried to understand what teachers know by developing methods to measure their knowledge and skills (e.g. Mertler & Campbell, 2005; Plake, 1993; Randel et al., 2011; Zhang & Burry-Stock, 1994). However, no research has been conducted with teachers in Brazil, especially with those teaching mathematics. Furthermore, as I explain next, my own experience has shown that research in this area is needed.

1.3 Prior research and the need to start from the beginning

Based on the episode presented at the beginning of this chapter and my recognition of the importance of assessment, during my MPhil I conducted an exploratory survey through an e-questionnaire seeking information about the current practices adopted by secondary school mathematics teachers in Brazil (Camargo & Ruthven, 2014; Camargo, 2015).

Teachers were asked, amongst other aspects, about the frequency of application, and the importance given to specific assessment methods or procedures. The results from the quantitative analysis showed that although the 332 respondents reported using various kinds of assessment and with different frequencies, tests and homework assignments were identified as the two most commonly used methods.

Additionally, although the literature argues that peer- and self-assessment are essential for improving learning, they were not common practice among respondents. In all questions addressing these aspects, teachers reported 'rarely or never' making use of them.

This research offered valuable information of 'what' teachers do, but did not provide further information of 'how' and 'why'. Therefore, in the first year of my PhD I went to Brazil with the purpose of better understanding teachers' practices to help them to start implementing FA in their classrooms.

Based on some interviews and observations that I conducted, I concluded that teachers were assessing their students without taking account of the varied elements that students' mathematical learning involves and that they had a narrow (if any) understanding of the foundational aspects of FA. I realised, therefore, that it would be necessary to step back and ‘start from the beginning’, collecting information on what mathematics teachers know about assessment and then proposing some changes.

Furthermore, just as in other parts of the world, there is little success, engagement or interest in mathematics in Brazil. Although analysis of international test data (OECD, 2014), shows improvements from year-to-year, it also emphasises a trend of systematic mathematics underachievement.

Considering that improving students’ learning, among other things, depends on how teachers operationalise the different dimensions of pedagogy (Muijs & Reynolds, 2010; Rowe, 2004), and bearing in mind that FA is one of these dimensions, the main goal of this study was to assess the FA knowledge of mathematics teachers in the FD of Brazil, through a structured online questionnaire developed for this purpose. The rationale for this was that the more that is known about teachers’ assessment knowledge and skills, the more it is possible to understand the kind of practices they can use in class. This information will be helpful in designing Continuing Professional Development (CPD) courses that fit their needs, target their weaknesses and develop their strengths even further.

1.4 Structure of the dissertation

The remainder of this dissertation is comprised of 10 chapters.

Chapter 2 sets the scene of this study by providing an overview of the education system in Brazil, highlighting the characteristics of the Federal District where this study was conducted.

Chapter 3 presents and discusses important elements related to FA by undertaking a review of the research literature as well as of the Brazilian and the FD official documents. The focus is to layout a framework of the essential aspects of FA that mathematics teachers in Brazil should know. Chapter 4 presents and explains the research questions and decisions with respect to the research design.

Chapters 5, 6 and 7 outline the initial steps taken for development of the instrument used for data collection, explore the qualitative and quantitative results of various pilot tests conducted to evaluate and validate the instrument, and detail the final version of instrument itself.

Chapter 8 explains the instrument delivery and the analysis of its results, followed by Chapter 9 which presents the findings of this enquiry, considering specific aspects of FD mathematics teachers' knowledge of FA and the behaviour of specific items when compared to the sample from the other states of Brazil.

Chapter 10 follows on from these results in an attempt to give answers to the research questions posed and to connect these proposed answers with other studies carried out in the field. This chapter also includes a discussion of the questionnaire developed during this enquiry as a new instrument to assess mathematics teachers' knowledge of FA.

Chapter 11 concludes this dissertation with a reflection on the implications and limitations of the study and suggests strands of research to be conducted in the future.

CHAPTER 2

THE BRAZILIAN CONTEXT

This chapter sets the scene for this study by presenting some characteristics of Brazil and its education system, highlighting some specificities of the FD where this study was conducted.

Section 2.1 offers some background information about Brazil and the FD. Section 2.2 presents an overview of the education system. Section 2.3 explains the teaching profession in Brazil and how its particularities in the FD compare with other states. Section 2.4 details how assessment takes place in its different dimensions.

2.1 Background

Brazil, officially the Federative Republic of Brazil, is situated in Latin America, and is the world's fifth-largest country, by both geographical area and population (approximately 208,608,900 inhabitants). It is the largest Portuguese-speaking country in the world, and the only one in the Americas. Figure 2.1 shows its territory with five geographic regions, 26 states and the FD.

The Federal Constitution and the General Law of Education 1996 (LDB-96)[1] determine that the Federal Government, States, FD and municipalities must manage and organise their respective education systems, with each system responsible for its own maintenance and funds. The constitution establishes a reserve of 25% from the state budget and 18% from federal and municipal taxes for education.

[1] Translated from 'Lei de Diretrizes e Bases da Educação' by the author.

Figure 2.1: Geographic regions and states in Brazil.

The FD is located in the Central-West region and has an area of 5,799,999 Km^2. Brasilia was specifically planned and developed to move the capital to a more central location in 1960, and as such, is the most important city of the FD. The name 'Brasilia' is commonly used as a synonym for the FD. However, the FD is composed of 31 administrative regions (Figure 2.2), with Brasilia being only one of them representing only 10% of the total population.

2.2 The education system

In the Brazilian education system, secondary schools[2] are free to all and compulsory for students between the ages of 11 and 14. The basic curriculum is compulsory

[2]Secondary schools are divided into Fundamental Education II, for students between the ages of 11 to 14 and Middle Education for students between the ages of 15 to 17.

Figure 2.2: Federal District and its 31 administrative regions.

and consists of: Portuguese, mathematics, history, arts, geography, science, physical education and a foreign language. Each region can supplement this core curriculum with other subjects as defined by the needs of the region or the school, and thus they become compulsory as well in that region. In the FD, for example, these subjects are taught for three lessons per week and are called 'diversified part'[3], with each school choosing the topic that will be studied.

Middle Education is also free, but is not mandatory. However, it is necessary to have completed this stage to go to university. Students are required to finish their Fundamental Education before they are allowed to enrol in Middle Education. The core curriculum is compulsory and comprises: Portuguese, two foreign languages, history, geography, mathematics, physics, chemistry, biology and physical education. Philosophy and sociology, banned during the military dictatorship (1964–1985), have since become compulsory again. Therefore, in contrast to the UK, Brazilian students do not choose their post-14 subjects and they are not associated with specific official qualifications. Every student will be taught all subjects and their contents will comprise the 'vestibular' (see Section 2.4.3).

Middle Education is also offered in the vocational sector, and the government (2003–2016) has invested a lot in this sector. The rationale is that since many Brazilian students do not have the opportunity to go to university, at least they will finish their

[3] Translated from 'parte diversificada' by the author.

studies with a profession, resulting in greater preparation for the labour market.

The school year, lasting 200 days (at least 800 hours of activities), is divided into four terms: usually beginning in February, with a 15-day break in July, and ending in December. Each shift[4] is divided into six lessons of different subjects usually lasting 50 minutes each. More weight is given to mathematics and Portuguese lessons, each with five lessons per week. The private sector also provides all levels, on a fee-paying basis.

Table 2.1 on the following page gives a better view of the divisions and sub-divisions of the Brazilian education system.

In the FD, the Department of Education is divided into 14 local authorities, which are responsible for administering the schools but are not responsible for their funding. Funding is sent directly to the schools from the Federal and local governments.

According to the last census, the FD is comprised of 905 schools, of which 63.2% are state schools.

2.3 The teaching profession

Until the LDB-96 was established, it was not necessary to have an undergraduate degree to become a teacher. Preparatory training for teachers was done during Middle Education as a vocational training stage called 'Magistério', which no longer exists. After LDB-96, the requirement to hold a university degree came into effect for all new teachers. Thus, anyone wanting to teach mathematics must have a degree in mathematics and have elected to do 'Licenciatura' as part of their degree, which encompasses three years of pure mathematics (calculus, topology, etc.) and one year of content related to teaching and learning mathematics in secondary schools.

Without doubt, the last year of the undergraduate degree alone is not enough to cover all issues related to classroom practice. A thorough study conducted by Gatti (2010), found that:

1. the proposed curriculum for teacher training courses has a fragmentary character, presenting a dispersed set of disciplines;

2. even among the specific disciplines, the approach is descriptive and there is almost no connection between theory and practice;

[4] Most schools have at least two shifts. In the school where I used to work, for example, it operated from 7am to 12 noon for Fundamental Education II, from 1pm to 6pm for Fundamental Education I, and from 7pm to 11pm for Youth and Adult Education.

Table 2.1: Divisions and sub-divisions of Brazilian education system.

Levels			Grades	Age
Basic Education	Regular	Early Childhood Education	Day Nurseries	0 to 3
			Kindergarten	4 to 5
		Fundamental Education I	1st	6
			2nd	7
			3rd	8
			4th	9
			5th	10
		Fundamental Education II	6th	11
			7th	12
			8th	13
			9th	14
		Middle Education	1st	15
			2nd	16
			3rd	17
		Professional	During Middle Education	15 to 17
			After Middle Education	Over 17
	Youth and Adult	Fundamental Education I	1st	Any age over 15
			2nd	
			3rd	
			4th	
			5th	
		Fundamental Education II	6th	
			7th	
			8th	
			9th	
		Middle Education	1st	Any age over 18
			2nd	
			3rd	
	Special	Exclusive Classrooms or Schools	Not divided into grades	Depends on the needs of each student
		Students included in regular classes	Follows the 'Regular' grades	

3. the content to be taught in secondary school appears only sporadically and, in most cases, is covered in a generic or superficial manner, suggesting weak association with teaching practices.

Therefore, the majority of teacher training is left to in-service courses, which are usually not compulsory and depend on teachers' interests. These courses vary according to each state but the federal government (2003–2016) has invested much effort in initial (for those teachers that have not graduated in the subject that they teach) and in-service teacher training programmes. Previous research, however, concluded that there is almost no evidence that recent teacher training courses have influenced mathematics teachers' approaches to assessment (Camargo, 2012).

On the other hand, teaching conditions in Brazil hinder teacher development. In most states, teachers have to teach three different shifts in order to earn a reasonable salary, meaning they do not have much time to prepare their lessons or undertake in-service courses. The exception is the FD, where teachers work 40 hours a week, divided into 25 hours of lessons in one shift and 15 hours of meetings and planning hours in the opposite shift. In this case, teachers can devote work time to planning lessons, talking to parents, marking homework and tests, etc., and undertaking courses. Furthermore, although still not adequate, the salary in the FD is the highest in Brazil[5].

Finally, to become a state-school teacher, one must pass a competitive test organised by each state according to its needs. Once approved, one becomes a teacher and the job is secure until one decides to retire or change jobs.

2.4 The assessment role

There are three major types of assessment that must be explained when referring to the education system in Brazil: classroom-based, external and the *vestibular*.

2.4.1 Classroom-based assessment

Classroom-based assessment is prepared, organised and applied by teachers each term during the school year. This can be done using different instruments or procedures. Some local governments determine the kind of assessment that must be made, whilst others leave it to teachers' choice. Each assessment generates a mark and the average mark across the year will decide if the student is able to progress to the next grade or not. This decision, therefore, will be made based solely on teacher-made assessments, and consequently have significant impact on teaching and learning.

In Brazil, although secondary mathematics teachers use various kinds of assessment and with varying frequencies, tests and homework assignments are the methods most

[5]At the FD, a teacher in his first year, receives around R$ 4,000.00 per month whereas in the other states, the salary is around R$ 1,500.00.

commonly used, both contributing towards students' marks (Albuquerque, 2012; Camargo & Ruthven, 2014; Camargo, 2015; Pacheco, 2007).

If students do not achieve the expected mark in one term, it can be 'recovered' in the next. If, by the end of the school year, students have still not achieved the expected pass mark, they have the chance to take a final exam (also teacher-made) to recover 'the mark'. If the pass mark is not achieved in more than three subjects by the end of the year, they must repeat the grade. If students do not achieve the mark in three subjects or less, the 'Partial progression with dependency system'[6] allows them to start the next school year in the next grade, but they will have to undertake work based on the curriculum of the previous year to recover the mark they did not achieve in the final exam. This can have many forms and it is the teacher who chooses what the student will have to do: a project, some exercises, some face-to-face lessons, a test, etc., or a combination.

2.4.2 External assessment

There is no face-to-face inspection in schools. Therefore, the ongoing diagnosis of the Brazilian education system is supported only by the external assessment system. The information produced can inform the design, redesign and monitoring of the education policies, contributing to the improvement in the quality, equity and efficiency of teaching. However, this does not always happen in practice.

External assessments are held every two years, when Portuguese and mathematics tests are sat by students. These are not high-stakes assessments as they have no influence on students' grades and no results are given to individual students. The Assessment System of Basic Education (SAEB)[7] is designed by the National Institute of Educational Studies and Research (INEP)[8] and consists of two large-scale evaluations:

National Assessment of Basic Education (ANEB)[9] covers a sample of students from public and private schools who are enrolled in the final grades of each stage of education: 5^{th} and 9^{th} grades of Fundamental Education and 3^{rd} grade of Middle Education.

[6]Translated from 'Regime de progressão parcial com dependência' by the author.
[7]Translated from 'Sistema de Avaliação da Educação Básica' by the author.
[8]Translated from 'Instituto Nacional de Estudos e Pesquisas Educacionais' by the author.
[9]Translated from 'Avaliação Nacional da Educação Básica' by the author.
[10]Translated from 'Avaliação Nacional do Rendimento Escolar' by the author.

National Assessment of Educational Achievement (ANRESC)[10] – also known as 'Prova Brasil' – is sat by students in the 5^{th} and 9^{th} grades of Fundamental Education only in public schools which have at least 20 students enrolled in the assessed grades. According to the School Census, it assesses students in approximately 60,000 schools, which comprise 86% of all enrolled students at this level of education.

Since 2007, data collected by the 'Prova Brasil' and the School Census are used to calculate the Index of Basic Education Development (IDEB)[11]. The IDEB is one of the most visible actions of the PDE – Plan for the development of Education[12]. This established a series of actions to be executed by the federal government, states and municipalities; targets for schools and municipalities to achieve within 15 years; and indicators for the verification of results. IDEB is a composite measure of the average of standardised test scores in reading and mathematics and the grade promotion rates at each stage of basic education. IDEB values vary from 0 to 10 and the goal is that all schools will have reached at least 6 by the end of 2022, which represents, in the Brazilian metric, the situation of a typical OECD country in 2007. A complete overview of the index can be found in Fernandes (2007) and a critical analysis in Soares and Xavier (2013).

Based on the IDEB, the quality of education in Brazil is improving, but not as planned. Table 2.2[13] shows that in 2005, the IDEB for Fundamental Education I was 3.8. The figure increased each year, and by 2013 it was 5.2, exceeding the target of 4.9 for that year. For Fundamental Education II, although it increased from 3.5 to 4.2 during the same period, it did not reach the 2013 target of 4.4. Similarly, in Middle Education it increased from 3.4 to 3.7, but did not reach the 2013 target of 3.9. The table also shows the difference between the results and the expected targets of private and state schools.

Table 2.3 shows that, in the FD, the situation although slightly better, is still not as expected, as the target was not reached in some years.

Schools are not punished for missing the targets, but receive more funding to improve their index. However, this funding cannot be used for hiring more staff, for example, or to pay training courses for teachers.

[11] Translated from 'Índice de Desenvolvimento da Educação Básica ' by the author.
[12] Translated from 'Plano de desenvolvimento da educação.' by the author.
[13] Available at http://ideb.in ep.gov.br/resultado/resultado/resultadoBrasil.seam?cid=262458.

Table 2.2: IDEB - Results and Targets

Schools	Results						Targets				
	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015
Fundamental Education I											
State	3.9	4.3	4.9	5.1	5.4	5.8	4.0	4.3	4.7	5.0	5.3
Private	5.9	6.0	6.4	6.5	6.7	6.8	6.0	6.3	6.6	6.8	7.0
Total	3.8	4.2	4.6	5.0	5.2	5.5	3.9	4.2	4.6	4.9	5.2
Fundamental Education II											
State	3.3	3.6	3.8	3.9	4.0	4.2	3.3	3.5	3.8	4.2	4.5
Private	5.8	5.8	5.9	6.0	5.9	6.1	5.8	6.0	6.2	6.5	6.8
Total	3.5	3.8	4.0	4.1	4.2	4.5	3.5	3.7	3.9	4.4	4.7
Middle Education											
State	3.0	3.2	3.4	3.4	3.4	3.5	3.1	3.2	3.3	3.6	3.9
Private	5.6	5.6	5.6	5.7	5.4	5.3	5.6	5.7	5.8	6.0	6.3
Total	3.4	3.5	3.6	3.7	3.7	3.7	3.4	3.5	3.7	3.9	4.3

Table 2.3: IDEB - Results and Targets - Federal District

Schools	Results						Targets				
	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015
Fundamental Education I											
State	4.4	4.8	5.4	5.4	5.6	5.6	4.5	4.8	5.2	5.5	5.8
Private	6.4	6.1	6.5	6.8	6.9	7.1	6.4	6.7	7.0	7.2	7.3
Total	4.8	5.0	5.6	5.7	5.9	6.0	4.9	5.2	5.6	5.8	6.1
Fundamental Education II											
State	3.3	3.5	3.9	3.9	3.8	4.0	3.3	3.4	3.7	4.1	4.5
Private	6.0	5.9	5.8	6.0	6.1	6.0	6.0	6.1	6.4	6.7	6.9
Total	3.8	4.0	4.4	4.4	4.4	4.5	3.9	4.0	4.3	4.7	5.1
Middle Education											
State	3.0	3.2	3.2	3.1	3.3	3.5	3.0	3.1	3.3	3.6	3.9
Private	5.9	5.5	5.6	5.6	5.7	5.6	5.9	6.0	6.1	6.3	6.6
Total	3.6	4.0	3.8	3.8	4.0	4.0	3.6	3.7	3.9	4.1	4.5

2.4.3 The *vestibular*

The third type of assessment is the ‘vestibular’, which is an exam students take at the end of their Middle Education if they wish to go to university. Each university used to have its own ‘vestibular’, but nowadays most use a common exam called National

Middle Education Exam (ENEM)[14]. This test is administered once a year and students can sit it as many times as they wish, which is a very common practice if a student wishes to attend a specific university and does not get the necessary grade in that year. It has a wide influence on teaching and learning during Middle Education for three main reasons:

1. The best universities in Brazil are public universities and therefore, everyone wants to take their courses. The fact that there is no tuition fee to be paid is also attractive.

2. There are not enough places for everyone in these universities, although places have increased substantially from 2003 to 2016.

3. Private schools usually use the number of students who successfully progress to university to advertise their own quality.

Because of the reasons above, in the last year of Middle Education, it is a very common practice to 'teach to the test' and some schools separate their students based on the university and/or degree they intend to sit the 'vestibular' for.

2.5 Summary

In this chapter, I have presented an overview of the Brazilian education system, explaining the different grades and the curriculum, as well as, what it means to be a teacher in Brazil. Emphasis was given to assessment because of the focus of this study.

Although all three kinds of assessment have major influence on the education system and on teaching and learning mathematics in secondary schools, only classroom-based was taken into account in this study, due to its relationship to FA. Further discussion can be found in the next chapter which discusses important issues related to assessment through a careful review of the literature.

[14]Translated from 'Exame Nacional do Ensino Médio' by the author.

CHAPTER 3

A FRAMEWORK FOR FORMATIVE ASSESSMENT

The previous chapters explained my motivation to conduct this study and outlined some specificities of the Brazilian education system. This chapter approaches FA in mathematics as the main concept of the research. The aim is to construct a theoretical framework by presenting and discussing the FA knowledge and skills that have already been considered in research literature (Bennett, 2011; Black & Wiliam, 1998; Heritage, 2007; Stiggins, 2007) and their connection to the teaching and learning of mathematics in secondary schools (Hodgen & Wiliam, 2006; Ma, 1999; Rakoczy, Harks, Klieme, Blum, & Hochweber, 2013; Wiliam, 1999). This chapter is split into two main parts.

The first part comprises six sections addressing essential aspects of FA that have been discussed in *research* literature. The second comprises two sections dedicated to reviewing the Brazilian and FD official documents to analyse to what degree the *institutional* literature encompasses the recommendations from the research literature. I start with an overview of what FA is.

3.1 What is formative assessment?

The complexity of defining FA has been discussed by various authors and institutions (ARG, 1999, 2002; Bennett, 2011; Black & Wiliam, 1998; Harlen, 2012). Although the term is usually attributed to Scriven (1967), much of the attention to FA started with

Black and Wiliam's (1998) seminal work.

At that time, Black and Wiliam (1998, p. 7) defined FA as encompassing "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged."

Since then, these authors and others have re-examined this definition and broadened the discussion in regards to what FA is and is not.

Popham (2008, p. 6) formulates his understanding of formative assessment as "a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics".

For Wiliam (2011, Chapter 2, Section 3, para. 22)

> An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted and used by teacher, learners or their peers to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have made in the absence of that evidence.

In Brazil, according to Villas Boas (2011), only in recent years has the term *formative assessment* appeared and its definition and principles are still not very well-known by teachers.

Some authors have suggested not using the term 'formative assessment' at all (e.g. Stiggins, 2005; Taras, 2005). Instead, they prefer to use the phrase 'Assessment for Learning (AfL)'. For others, (e.g. Swaffield, 2011, p. 443), there are important differences between both that need to be understood for their appropriate use:

- AfL is a learning and teaching process, while FA is a purpose and some argue a function of certain assessments;
- AfL is concerned with the immediate and near future, while FA can have a very long time span;
- The protagonists and beneficiaries of AfL are the particular pupils and teacher in the specific classroom (or learning environment), while FA can involve and be of use to other teachers, pupils and other people in different settings;

- In AfL pupils exercise agency and autonomy, while in FA they can be passive recipients of teachers' decisions and actions;
- AfL is a learning process in itself, while FA provides information to guide future learning; and
- AfL is concerned with learning how to learn as well as specific learning intentions, while FA concentrates on curriculum objectives.

Although the definition has been re-examined by many, and there are some points where they have not reached a consensus, they all seem to agree, even if implicitly, that FA is not a type of assessment, but a process in which assessment information is used – by teachers and students – to improve learning; and should be considered formative only if information is indeed used for that purpose. Replacing the term 'formative assessment' with 'assessment for learning' does not help in clarifying the issue with the definition (Bennett, 2011). As put by Good (2011, p. 4) "the challenge ahead of us is to put into practice the presumption that the label applied to an assessment is far less important than what is done with the information gathered."

Sharing Bennett's (2011) and Good's (2011) views, in this study the terms 'formative assessment' and 'assessment for learning'will be used indistinctly and interpreted as a process in which both teachers and students actively become the agents of the process, willing and responsible for their own knowledge and practice. In this process, *formally* gathered evidence is used to formulate feedback and inform decisions; and *informal* evidence (e.g. observation, conversations) is used to generate teacher and peer-feedback to improve learning (Hargreaves, 2005; Wiliam & Leahy, 2015).

However, to be able to understand what FA looks like in practice, and which knowledge and skills teachers need to be able to put FA into practice, it is necessary to expand this definition. This is the purpose of the next sections.

3.2 The framework

As said before, the purpose of the following sections is to present a framework that expands on the knowledge and skills teachers need to be able to put FA into practice. As such, it should be situated in a broader framework of professional knowledge as discussed by several authors (Bromme & Tillema, 1995; Leinhardt, Young, & Merriman, 1995).

Bromme and Tillema (1995) define professional knowledge as the activity-oriented knowledge of practitioners: this includes information about situations and methods of problem solving and also information required to define and understand the problems a professional faces.

For Leinhardt et al. (1995), from a cognitive point of view, professional knowledge is resultant of professional action, through the integration, tuning and restructuring of theoretical knowledge to the demands of practical situations and constraints. From a socio-historical perspective it evolves gradually guided by the culture of the professional within a working context which is in itself part of a certain culture. What these approaches have in common is that professional knowledge is built through practice and experience but makes use of theoretical knowledge.

More specifically, in relation to teachers' professional knowledge, there is a body of studies that emphasise teacher-based knowledge and experience over university research-based knowledge and empirical theory (Carr, 1992; Fenstermacher, 1994; Schön, 1992). Their argument refers to academic knowledge being too abstract and general, while teaching is concrete and specific; or that it makes teachers subservient to the producers (researchers, academics) of that knowledge and alienated from their jobs.

On the other hand, some authors advocate that theoretical knowledge is required for practical professional activity, but to provide a basis for practical knowledge, theory requires transformation (Bromme & Tillema, 1995). Their argument is that although theory can afford deep insight, explanation, and understanding, it will never be able to represent the reality in full (Norris, 2000). Teachers mediate research-based knowledge and their classroom situations. That is, through their experience, reflection on their practices, and professional judgement based on their knowledge of their students and their classrooms, teachers' 'abstract and general' theoretical knowledge is transformed into knowledge which is 'concrete and specific' to a teaching situation.

In this study, professional knowledge is not approached as a theory-practice dilemma but as integrated entities complementing each other. As said by Bromme and Tillema (1995, p. 266), "becoming a professional is not a process of substituting theory by experience, but a process of fusing theory and experience". Teachers will not be successful in implementing FA by only relying on their theoretical knowledge (learned in academia or on CPD courses), but equally will not be able to implement FA and improve their practical knowledge if they do not have a strong knowledge base and are not constantly revisiting it.

Therefore, this study considers both types of knowledge. The framework discussed in the following sections is developed based on the theoretical or academic research-based knowledge of FA. Chapters 5, 6 and 7 explain how this knowledge was operationalised into a structured questionnaire with situations involved in the professional work of FA, considering the teacher-based practical experiential knowledge and its relation to classroom practice (Grossman, 1995).

With that said, it is important to highlight that the framework has two characteristics that need to be explained. The first is that it is designed *for* mathematics teachers, however, there will be no specific section that approaches FA only related to mathematics. Its characteristics will be embedded and exemplified within all parts of the framework. Additionally, although several parts of the framework can also be applied to other subjects, the examples and explanations will always be related to mathematics.

The second is that the framework takes feedback as the heart of FA. It considers feedback as the core idea; that it should not be considered just a component of FA. Rather, true FA is not occurring in the absence of feedback and assessing students with formative purposes all flows from the understanding of the feedback process. Thus, the framework discusses essential issues about how mathematics teachers should translate ideas about the feedback process into practical actions that they take when formatively assessing their students.

As feedback is the key driver of the framework, I start by explaining its importance.

3.2.1 The importance of feedback

Feedback has been defined by many authors. According to Sadler (1989, 2010), feedback can be seen as teachers giving information to students on how successful they were in the development of their work, and helping students to identify what is missing and how to close this gap. Hattie and Timperley (2007) developed a model for effective feedback that incorporates learning goals, progress towards those goals, and the steps needed to make better progress towards those goals. It consists of helping students to answer three crucial questions:

1. Where am I going? (What are the goals?)
2. How am I going? (What progress has been or is being made towards the goal?)
3. Where to next? (What activities need to be undertaken to make further progress?)

Many studies have provided sufficient evidence that effective feedback helps with the improvement of learning (Bee & Kaur, 2014; Black & Wiliam, 1998; Crooks, 1988; Darling-Hammond, 2015; Kearney, Webb, Goldhorn, & Peters, 2013; Kluger & DeNisi, 1996; Hattie & Timperley, 2007; Natriello, 1987). Effective feedback can also lead to increased autonomy, critical thinking, character building, and improved behaviour (Newman, 2016).

Feedback can be helpful, for example, in enabling students to plan their course of action in their learning of mathematics (Bee & Kaur, 2014). Similarly, constructive feedback on students' homework assignments not only enhances achievement (Núñez et al., 2015), but also helps to minimise the differences between boys and girls and improve attitudes towards mathematics (Elawar & Corno, 1985).

On the other hand, feedback based on 'comments only' produces more learning gains than 'marks only' or 'comments with marks' (Butler, 1988). Feedback given as grades has also been shown to have especially negative effects on the self-esteem of low-ability students (Craven, Marsh, & Debus, 1991), and on achievement and motivational variables (Lipnevich & Smith, 2009). Similarly, feedback that draws attention away from the task and towards self-esteem can have a negative effect on attitudes and performance (Black & Wiliam, 1998; Hattie & Timperley, 2007).

Therefore, what these studies have also shown is that not all feedback is equally effective, which indicates that students' learning will not improve with the mere provision of feedback (Owen, 2016). Consequently, teachers should know that there are different kinds of feedback and that each should be used in different situations and for different purposes.

Hattie and Timperley (2007) argue that the effectiveness of feedback will depend on the level at which it is aimed. They propose a model of four levels of feedback:

Feedback about the task: related to how well a task is being accomplished or performed, such as distinguishing correct from incorrect answers, acquiring more or different information, and building more surface knowledge, such as "You correctly used the tables to find out the price of the shirts. What can you do to find out the cost of buying the packages for delivering the shirts?"[1] It can also be related to neatness and task accomplishment, such as "You didn't answer the second part of the question. How would you know if you had all possible combinations?" (Brookhart, 2008, p. 93).

[1] Adapted from (RAPPS, n.d., p. 3).

However, feedback at task level has a drawback: as it is specific to a particular task, it may not be transferable to other tasks or situations. In this case, feedback about the processing of the task may be more effective (Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002).

Feedback about the processing of the task: related to the process used to understand and solve the task, such as “I can see that you understood what Jill meant by doubling the height and width. Your diagram correctly shows what would happen if you doubled each side of the garden. Your next step is to find the areas for each garden in your diagram. Why don’t you look for the word ‘area’ in your mathematics journal?"[2]

Feedback about the processing of the task can also help students to develop strategies for error detection. Therefore, it can lead to more effective information-search and the use of task strategies, also influencing students’ self-regulation.

Feedback about self-regulation: related to students’ self-evaluation or self-efficacy in engaging on the task, such as “I see you figured out that this problem asks you to make a list of combinations of ice cream flavours and containers [...]” (Brookhart, 2008, p. 93).

This type of feedback is closely related to motivational aspects: whether the student is willing to invest effort to deal with the feedback (Marsh, 1990), the degree of confidence or certainty about the correctness of the response, and the attributions about success or failure (Salili & Hau, 1994), and therefore to their capability of self-assessment.

Feedback about the self as person: it is personal, in the format of praise, and not related to the task, such as “You are so clever”, “Good girl” and “Well done”.

As this kind of feedback does not offer information about ‘where they are’ and ‘what’s next’, it does not lead to learning gains (Kluger & DeNisi, 1998). As noted by Stiggins (2009, p. 239) “communication supports learning when it focuses on attributes of the student’s *work* and not on attributes of the student.” [my emphasis]

[2] Adapted from (RAPPS, n.d., p. 3).

> However, if praise is directed to students' effort, engagement and the processing of the task, such as "You are an excellent student because you figured out that this problem asks you to make a list of combinations of ice cream flavours and containers."[3], it can help to increase self-efficacy and can, therefore, impact on the task.

Although the level at which feedback is aimed is important, there are other issues that teachers should take into account, which I approach next.

3.2.1.1 Timing

The timing of feedback is also of great importance. Although some studies have shown that immediate feedback is most effective (Corbett & Anderson, 2001; Epstein et al., 2010) and therefore it should be delivered as soon as possible (Chan, Konrad, Gonzalez, Peters, & Ressa, 2014), it should be given too early, before students were able to give sufficient thought to the task.

For example, Simmons and Cope (1993) concluded that immediate feedback given to students using Logo inhibited moves to a higher level of response compared to those levels of response reached when responding to paper tests. They believe this was because students using Logo were taking advantage of the immediate feedback, and solving the task by trial-and-error attempts, and therefore with little mental effort. As with the paper and pencil solution, trial-and-error was much more time consuming, students discovered that it would be more worthwhile if they thought more carefully about the solution, which in turn, resulted in more learning gains. However, it could also be argued that, in this case, the feedback is integral to the problem solving task environment, and not just because it is immediate.

On the other hand, feedback timing will also depend on the level at which feedback is aimed (Hattie & Timperley, 2007), and most of the research has been conducted without considering them. At the task level, immediate feedback is beneficial, whereas at the processing level, some delay may be required as students may need more time. In addition, the teacher may need time to go over students' work, outside the classroom, in order to formulate useful comments and feedback, either in writing or as a plan for a future conversation with students (Laveault & Allal, 2016).

Equally, timing of feedback is also related to the difficulty of the task. Clariana (1999), for example, concluded that delayed feedback is beneficial for difficult tasks, as

[3] Adapted from (Brookhart, 2008, p. 93).

they usually require a greater degree of 'processing about the task' (Hattie & Timperley, 2007).

Therefore, the relationship of feedback timing and students' learning will vary based on the type of the task and at which level it is aimed. However, in any case, it should be delivered while the assignment or learning task is still relevant to the student (Brookhart, 2008), so students will be able to act upon the feedback while they are still aware of the learning target and while there is still time for them to act on it.

3.2.1.2 Content

Concurrently with the importance of timing in the process of feedback, teachers should also be careful about the information they provide to students. They need to understand that giving feedback on mathematical learning should go beyond indicating an incorrect answer, because seeing an answer as simply wrong will not equip a teacher with the detailed mathematical understanding required for a deeper treatment of the problems that students are facing (Ball, Thames, & Phelps, 2008).

Therefore, students' mistakes can no longer be considered a failure liable to sanctions, but rather as an essential source of information, whose manifestation it is important to encourage. Consequently, the classroom atmosphere needs to reflect this view that mistakes are an opportunity to learn, so that critical feedback is perceived as constructive instead of judgmental (Heitink, der Kleij, Veldkamp, Schildkamp, & Kippers, 2016). Students must trust that feedback from the teacher is intended to move the learning forward, rather than to point out failure.

Feedback improves performance when its content is focused on what needs to be done to improve, and particularly when it gives specific details about how to improve (Hoffmann, 2005; Wiliam, 1999). It should be worded in such a way that it helps the receiver correct inappropriate task strategies, procedural errors or misconceptions (Mason & Bruning, 2001). Thus, effective feedback should provide a recipe for future action. Feedback like "you need to work more on that" or "you need to study more" does not show students how to improve, and therefore is highly unlikely to improve learning. Feedback needs to be related to goals, standards or criteria (Nicol & Macfarlane-Dick, 2006), which I discuss further in Section 3.2.3.

Sadler (2010, p. 538) agrees when he states that feedback "has to be both specific (referring, as it necessarily does, to the work just appraised) and general (identifying a broader principle that could be applied to later works)". Therefore, feedback should start by noticing what the student did, and then pointing out what they missed and how they can improve, such as "This is a well-developed proof, but I think you might have mixed up the definitions of rational and irrational numbers. If you change those in your argument, you will be on a great track." (Bleiler, Thompson, & Krajčevski, 2014, p. 118).

Brookhart (2008) suggests that feedback in mathematics should have the problem solving process as reference, and therefore should point out whether or not the student:

- identified the problem (What exactly is being asked here?);
- identified the elements of the problem (e.g. Which numbers need to be added? Which numbers are irrelevant?);
- identified one or more strategies that would work to solve the problem and then choose and apply one of those strategies;
- evaluated the solution they arrived at to see if it is a reasonable response to the problem.

For example: "I notice that you list the flavours in order – vanilla, chocolate, and strawberry – twice. Keeping the list organised is a good way to make sure you don't miss any. But there are 9 combinations, and you found only 6. Can you figure out a way to organise both the flavours and the containers at the same time to get all 9?" (Brookhart, 2008, p. 93).

By pointing out which of these steps the student has achieved ("I notice that you list the flavours in order"), teachers will be giving feedback about self-regulation (Hattie & Timperley, 2007), which is related to students' feelings of self-efficacy and has been shown to help with students' engagement and performance on tasks (Barry, 2008; Craven et al., 1991; Harks, Rakoczy, Hattie, Besser, & Klieme, 2014; Kluger & DeNisi, 1998).

On the other hand, indicating or giving cues about which of the steps they missed or misunderstood ("But there are 9 combinations, and you found only 6") – and how they can close the gap ("Can you figure out a way to organise both the flavours and the containers at the same time to get all 9?"), is more related to feedback about the task and about the processing of the task, which is also important for the improvement of students' learning.

The amount of information is equally important. As feedback has been shown to be so important for teaching and learning, it is natural to think that it should be given in large amounts. However, feedback works most effectively when it arrives in amounts the student can process versus in amounts so large as to overwhelm (Stiggins, 2009). At the same time, “what makes the difference is a usable amount of information that connects with something students already know and takes them from that point to the next level” (Brookhart, 2008, p. 12). Therefore, students should get enough feedback so that they understand what to do, but not so much that the work has been done for them. Overly directive, dishonest, wordy and unfocused feedback is perceived as irrelevant and frustrating by students both in primary and secondary levels (Hargreaves, 2013; Kay & Knaack, 2009), and in higher education (Lizzio & Wilson, 2008).

Day and Cordón (1993), for example, showed that students who received feedback in the format of ‘scaffolded’ help (they got only the minimum amount of help necessary to be able to move on with that activity), showed greater improvement then those who received a ‘full’ answer as soon as they found a difficulty in the task, and then were given another task.

However, in any situation, feedback will not lead to improvement until students understand both the feedback itself and how it applies to their work. Hence, teachers need to understand that the information about the distance between students’ current level and the expected level can only be considered as feedback if it is used to change this distance (Black & Wiliam, 1998). If the information is simply recorded and delivered to people not directly involved in the situation (or who do not have the power to change it); or made so coded that it is almost impossible to implement an appropriate action, such data will become useless. Therefore, the relationship between feedback information and how students receive it, is also very important in the feedback process.

3.2.1.3 Students and the feedback process

Complementary attention should be directed to what students make of the feedback, rather than just its timing or composition (Sadler, 2010). First, the only way to tell if learning indeed resulted from feedback, is allowing (and encouraging) students to make some kind of response to complete the feedback loop (Sadler, 1989; Nicol & Macfarlane-Dick, 2006; Reinholz, 2016; Whitington, Glover, & Harley, 2004).

Second, feedback can influence how students feel about themselves (whether they can see themselves as capable or not to achieve the expected target) (Brookhart, 1997;

Garcia, 1995; Stiggins, 2009), and what and how they learn (Covington & Omelich, 1984; Dweck, 1999; Stiggins, 2009). Therefore, feedback can regulate and be regulated by motivational beliefs (Brookhart, 1997; Nicol & Macfarlane-Dick, 2006).

What it is important to understand is that different learners respond better to different types of feedback. Brookhart (2008), for example, suggests that for 'successful students', teachers should focus on the task and the process, and be criterion-referenced, whereas for 'struggling students', self-referenced is better, since it shows a target that can be reached. In other words, for struggling students, feedback should be based on their past performance or on teachers' expectations based on this previous performance.

This is because, in some cases, struggling students are struggling simply because they do not see themselves as capable of learning. If the teacher shows that they actually improved (based on past performance), it can help them to verify their sense of efficacy for learning, as a guide or check for their own self-assessment. This will help them learn how to learn and achievement will grow (Hattie & Jaeger, 1998; Stiggins, 2009).

Similarly, if students do not see their learning as something that can be improved, feedback will lose its function. What teachers need to understand, is that feedback given to students should support a view that ability can be improved, and therefore is not fixed as an attribute of the student (Wiliam, 1999). Dohrn and Bryan (1994) discussed studies that have shown that providing positive ability attributions, in combination with student instruction in specific task strategies (e.g., mathematics and reading), leads children to persist longer and acquire adaptive attributions.

In other words, teachers cannot force students to focus on or learn something. However, when teachers take the time to provide feedback, students are more likely to be engaged in their class work (Kearney et al., 2013). This is particularly relevant in mathematics classrooms where concepts build on one another. Teacher feedback is input that, together with students' own internal input, will help students decide where they are with regards to the learning goals they need to meet, and what they should focus on next (Brookhart, 2008).

However, if productive feedback is the one which informs students how to do better next time, it needs to be fed into a classroom environment where there will be a next time (Stiggins, 2009). That is to say, "unless students are able to use the feedback to produce improved work, through for example, re-doing the same assignment, neither they nor those giving the feedback will know that it has been effective" (Boud, 2000, p. 158).

Nicol and Macfarlane-Dick (2006) suggest that feedback should be conceptualised as dialogue, and not as information transmission, meaning that students should have the opportunity to engage with the teacher in discussion about that feedback and not only passively receive it. For them, these discussions are crucil for students "to develop their understanding of expectations and standards, to check out and correct misunderstandings and to get a response to difficulties" (p. 210).

According to Sadler (1989), for students to be able to use teachers' feedback to close the gap in their learning, they must already have some assessment skills like their teachers. Thus, it is important that feedback is given to develop the skills for students to be able to assess the quality of their work by themselves.

On the other hand, teachers need to understand that feedback does not need to come from the teacher only, that the transition from teacher feedback to students' self-regulation does not happen automatically, and that it will not be developed satisfactorily simply by talking to students about them (Sadler, 2010). The development of this assessment capacity is part of the learning to be acquired by the student.

Based on that, for the feedback process to function properly, teachers should put in more effort and spend more time on developing self- and peer-assessment skills in their students (ARG, 1999, 2002; Black & Wiliam, 1998; Crooks, 1988; Yorke, 2003). In the next section, I discuss the importance of both practices.

3.2.2 Self- and peer-assessment

As previously noted, it is essential that students have an active role in the feedback process. In fact, the ultimate goal in the process of feedback is to help students develop their own skills to assess their work, and therefore to be able to detect the gap between the situation of their actual learning (How am I going?), the desired goal to achieve (Where am I going?), and what they can do to remove this gap and achieve the goal (Where to next?). Many practices can be adopted to help students develop these assessment skills, but two of them are the most widely discussed in the literature: self- and peer-assessment.

"Self-assessment means that students make judgements about their own achievement and learning processes and take part in decisions about action for further progress in learning", and "peer-assessment involves students in assessing each other's work, through reflection on the goals and what it means to achieve them" (Sebba et al., 2008, p. 6).

Thus, self-assessment must refer to the process by which students themselves, with the help of their teachers, continuously analyse the activities they have developed and those they are still developing, record their perceptions and feelings, and identify future actions to make progress in their learning; whereas in peer-assessment, this is done by a peer. This analysis must take into account: what they have and have not yet learnt; what was easily done; and which activities or concepts were difficult to understand and/or solve.

However, some students still think that assessment is a teacher's responsibility (Cowie, 2005; Weeden & Winter, 1999), and this can be happening because teachers are not preparing their students (or giving opportunities to them) to undertake this role. Black and Wiliam (1998, p. 25), for example, state that a "focus on self-assessment by students is not common practice, even amongst those teachers who take assessment seriously". Earl and Katz (2008, p. 91) add that: "Teachers rarely think pro-actively about what they need to do to use assessment to promote student self-assessment and self-regulation so that students become adept at defining their own learning goals and monitoring their progress towards them".

Self- and peer-assessment are also not common practice among teachers in Brazil (Camargo & Ruthven, 2014). When used, it is only with summative purposes. Villas Boas (2011), for example, observed that teachers are used to preparing a script to be answered by students, at times chosen by teachers, usually at the end of a period or an activity, and frequently asking students to assign a grade for it. She also observed that students usually do not know what this kind of self-assessment is for and what will be done with the information provided. Most students were assigning grades only thinking about the grade, which was therefore having no effect on their learning.

Therefore, teachers need to first understand the purpose of self-assessment, which cannot be for students to assign themselves grades. It should have the main purpose of enabling students to continuously reflect on their learning and develop the ability to record their perceptions. If teachers change their focus, students may change too.

This was the case in Klenowski's (1995) study, for example. The author found that when students were participating in self-assessment, the process of evaluation (including detailed feedback from the teacher) was more important than the grade itself. These students also described how they ended up being more honest about their abilities and their own work. For (Brookhart, 2008), self-assessment increases students' interest in feedback because it is 'theirs'; it answers their own questions and helps them to develop the self-regulation skills necessary for using any feedback.

This has also been the conclusion of other studies. In the study of Brookhart, Andolina, Zuza, and Furman (2004), despite having a more summative approach, students were invited to predict their scores and then reflect on their results in multiplication fact tests for ten weeks. The authors concluded that participating in the reflection helped students articulate the value of their own studying and also to develop strategies to increase the scores during the following week.

In a more general and qualitative approach, teachers involved in the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP) (Black, Harrison., Lee, Marshall, & Wiliam, 2003) used the 'traffic lights' idea to start using self-assessment in their classrooms and encourage students to think about their work. Students should catalogue their activities according to the three different colours: green for those concepts and/or activities which they did not have any problem solving or understanding, amber for those which were not completely clear to them and red for those with which they had serious problems in comprehending.

Although it was a simple idea, teachers reported that the implementation of self-assessment was not easy at the beginning because "their students lacked the necessary skills both to judge specific problems in understanding and to set realistic targets to remedy problems within reasonable time frames" (Black et al., 2003, p. 50).

In fact, for self-assessment to truly become an important component of students' involvement in assessment, it should work closely together with teacher feedback. Taras (2003), for example, has compared students' self-assessments in two different situations: prior to teacher feedback and after receiving teacher feedback. The results showed that both conditions benefited learning. However, the latter helped students identify and correct more errors (those that they or their peers had not been aware of) than in the first situation.

Therefore, as argued in Section 3.2.1, to have a well-functioning feedback process, teachers should also provide opportunities for students to develop the skills necessary to better judge and analyse their own learning and to start thinking about what actions they should take to improve. This includes a classroom environment in which students can work together and talk about their work.

Yackel and Cobb (1996) made the important observation that the daily practices and rituals of the classroom play an important part in how students perceive and learn mathematics. Cobb, Wood, and Yackel (1993) reported that students create 'insider' knowledge of mathematical behaviour and discourse from the norms associated with those daily practices. This knowledge evolves as students take part in the "socially developed and patterned ways" (Scribner & Cole, 1981, p. 236) of the classroom.

By scaffolding the development of those patterned ways, the teacher regulates the mathematical opportunities available in the classroom.

That is to say that teachers need to be committed to learners having control over the process (O'Shea, 2015), and to be able to discuss learning and develop effective student feedback, because classroom culture is related to positive outcomes for students (Sebba et al., 2008). Self- and peer-assessment are more likely to impact on student outcomes when there is a move from a dependent to an interdependent relationship between teacher and students, which enables teachers to adjust their teaching in response to students' feedback. Therefore, peer-assessment also plays an important role in the assessment process and may even be the first step towards self-assessment.

There are many reasons that justify the implementation of peer-assessment and the benefits of its implementation as a complement to self-assessment. First, when students know that their activities will be appreciated by their colleagues, they tend to prepare them more carefully and possibly with more pleasure. Secondly, when students work in pairs or in groups of three or four, they feel more comfortable just because they can conduct the discussion using the language that they normally use. Third, while students are working together, the teacher has more time to help those who need more support. Fourth, when students participate in peer-assessment, they can practise how to apply the criteria for good work, and how to value feedback (Black et al., 2003; Brookhart, 2008).

However, as well as with self-assessment, what is required is not peer-assessment as routine activity or as a means to keep students 'busy', but purposeful peer-assessment that is designed with a clear pedagogical focus, with the intention to provide students with practical experience to provide and receive feedback, consequently improving their mathematical learning (Boud, Cohen, & Sampson, 2001; Sadler, 2010).

Finally, teachers need to understand that like self-assessment, peer-assessment also needs to be supervised and supported by them and this can take time and practice (Black et al., 2003; Brookhart et al., 2004). A well-known barrier is students' lack of understanding of the learning intentions and the success criteria, and some have suggested that students should be involved in co-designing them (Sebba et al., 2008). In the next section, I address the importance of the learning intentions and success criteria to the feedback process and consequently in FA.

3.2.3 Learning intentions and success criteria

Clarifying, sharing and understanding learning intentions and success criteria (or transparency (Frederiksen & Collins, 1989, p. 30)) are very important steps of teaching and learning, and consequently of classroom assessment. They are related to whether teachers know how to make clear to their students what is expected from them (Where am I going?), but also whether students understand what they are meant to be doing (Wiliam, 2011), which is essential in the feedback process. They are important both for teachers and students.

Teachers will use them as the basis of their work. The learning intentions will be the starting point of their lessons. It is through them that teachers will decide which tasks they should give to their students, and which methods they will use to collect evidence of students' learning. After that, teachers will compare what students produced with the learning intentions. Having clear learning targets is the first step in the feedback process (Black & Wiliam, 1998; Brookhart, 2011; Hattie & Timperley, 2007; Sadler, 2010; Shepard, 2002; Shute, 2008). Teachers will only be able to analyse whether their students achieved the expected goals, and consequently feed back to their students effectively, if teachers themselves know what they want from their students.

More than that, these intentions must be shared with their students (Stiggins, 2009; Wiliam, 2011). If students are aware of the targets they are trying to achieve (Where am I going?), it will enable them to know whether or not they have achieved what was expected, developing a better grasp of their own strengths and weaknesses (How am I going?).

For that reason, specific learning intentions are more effective than general or non-specific ones (Brookhart, 2008; Hattie & Timperley, 2007; Shute, 2008; Wiliam, 2011). It will facilitate the teacher to include information on how to achieve them (the success criteria) making the feedback more directed (Latham & Locke, 1979; Chan et al., 2014) and more likely to get students' attention.

However, the experience of the King's Oxfordshire Summative Assessment Project (KOSAP) (Black, Harrison, Hodgen, Marshall, & Serret, 2011) has shown that the clarity of presentation of learning intentions, and also the activities, assignments, and assessments is a necessary condition, but not wholly sufficient.

Students' involvement in the feedback process will depend not only on how teachers communicate the learning intentions or success criteria, but also how students appropriate them. The meaning that students assign to what the teacher proposes is very important. Under these circumstances, teachers not only need to make them clear to students, but they also need to have strategies to check if students understand what is expected from them (Wiliam, 2011).

The more important reason for helping students develop an understanding of learning intentions and success criteria is to directly improve learning and to develop meta-cognitive knowledge for monitoring their own efforts (Shepard, 2002). Newby and Winterbottom (2011) found that providing assessment criteria was also important for the success of peer- and self-assessment.

Equally, if students do not understand what is expected from them, they will not be able to act upon teacher's feedback and consequently, they will not be able to improve their learning. Raising students' understanding of the assessment criteria is essential to initiate the productive dialogues which need to occur in the feedback process (Rust, Margaret & O'Donovan, 2003). For that, teachers need to be able to articulate these intentions in a way that is meaningful, challenging and attainable to students and also assessable (Brookhart, 2011; Hattie, 2009).

Furthermore, teachers need to understand that students' appropriateness of the learning intentions and the engagement on classroom tasks are not just related to cognition, but also to motivational beliefs. Goals are more effective when students share a commitment to attaining them, because they are more likely to seek and receive feedback (Latham & Locke, 1979).

> The mere provision of explicit criteria will not enable learning in all the ways desired if they are imposed autocratically and mechanically applied. [...] Students have to have the opportunity to learn what criteria mean (surely not memorize them as a list), be able to apply them to their work, and even be able to challenge the rules when they chafe (Shepard, 2002, p. 61).

The idea is to analyse whether the current 'modes of engagement' that teachers are using are working, or if some type of change is necessary (Nicol & Macfarlane-Dick, 2006). Wiliam (2011) proposes, for example, that teachers should bring examples of students' work to be analysed with students in class. These 'exemplars of performance' are effective because they make explicit what is required, and they define a valid standard against which students can compare their work (Orsmond, Merry, & Reiling, 2002). Another strategy would be a collaboration in which teachers and students negotiate and establish the success criteria for a piece of work.

At the same time, teachers need to understand that rarely will one method of assessment be suitable to assess all the learning intentions at the same time. A broader range of assessment tools is needed to capture important learning goals and to more directly connect assessment to ongoing instruction.

3.2.4 Using different methods for collecting evidence of students' learning

The use of different methods for assessing students is encouraged by many authors for several reasons. First because, as said previously, rarely will one method of assessment be suitable to assess different learning intentions at the same time.

Therefore, teachers should understand the purposes and uses of a range of available assessment options and be skilled in using them (Brookhart, 2011). In other words, teachers should understand that there are several options which include: tests, observations, seat-work, homework, oral questioning, self- and peer-assessment, portfolios, questionnaires, projects and various other methods. This is in addition to understanding that each method serves different purposes, can be incompatible with their instructional goals and consequently may impact quite differently on their teaching (AFT, NCME, & NEA, 1990) and on the process of giving feedback to their students.

If the goal is to help students learn, the methods used should be suitable for students' needs (McMillan, 2000; Popham, 2009; Suskie, 2002), which is the second reason for using different methods of assessment. Therefore, by developing and/or choosing different methods to assess their students, teachers should take into account that students learn and demonstrate their learning in various ways. Consequently, teachers also need to include opportunities for students to show what they have learnt using different forms of presentations (oral, written or visual).

Otherwise, reducing the assessment of mathematics to the students' written production is to reduce their potential of doing mathematics to their ability to write it down. This fact denies that the mathematical activity, before being a written production, occurs in terms of ideas, thoughts, and intuition. The ideas are, at first, mentally represented in a symbolic basis, which is not necessarily the same system used for mathematical writing. The mathematical thinking is much broader and sometimes more powerful than the one portrayed in the written form.

Therefore, when thinking about developing and/or choosing the method for collecting evidence of students' learning, teachers must somehow be able to cause the student to externalise the learning that has occurred, so teachers can use the information as an investigative tool to find out as much as they can about what their students know and can do, and what confusions, preconceptions, or gaps they might have. It should be designed to make each student's understanding visible, so that teachers can decide what they can do to help students to progress (Earl & Katz, 2008), and it requires the organisation of experiences, situations and contexts that allow students to show what they have learnt.

In the study of Black et al. (2003), for example, when teachers had to create useful feedback comments for their students, they realised that some tasks they had set needed to be reassessed, mainly because the tasks were not providing good information regarding their students' progress.

Students need to be given the conditions to communicate and argue the mathematical procedures they developed and used to get to a certain result (be it right or wrong), which includes well-developed assessment tasks as well as a class environment which encourages students to use assessment information for their own learning (ARG, 2002; Wiliam, 2011).

Finally, as each method has limitations, by using different methods it is easier to get a complete picture of students' understanding and enables teachers to make more valid inferences with fewer errors.

This is especially important in Brazil as most teachers still rely on teacher-made tests and homework to assess their students (Albuquerque, 2012; Camargo & Ruthven, 2014) and students' futures will be decided only based on teacher-made assessments. Therefore, it is very important for teachers to be able to rely on these results to make sound decisions about instruction and, consequently, to provide effective feedback to their students.

In addition, it is a common practice to seek ideas from textbooks or instructional materials. Therefore, it is very important that teachers know how to analyse these materials because they were not primarily intended to assess students' learning, and therefore might not be suitable for this purpose or might not be compatible with the content developed during the classes.

In other words, teachers need to make sure that the questions or activities they chose to assess students' learning are in agreement with the intentions that were previously made clear to them. Otherwise, this will jeopardise the feedback process, students' self-assessment, and consequently students' motivation for learning. Teachers need to understand how to create and use activities, assignments and assessments that embody the learning intentions at different levels of students' understanding (Ruiz-Primo, Furtak, Ayala, Yin, & Shavelson, 2010).

As Earl and Katz (2006, p. 57) comment:

> [...] the methods chosen need to address the intended curriculum outcomes and the continuum of learning that is required to reach the outcomes. The methods must allow all students to show their understanding and produce sufficient information to support credible and defensible statements about the nature and quality of their learning, so that others can use the results in appropriate ways.

Finally, as it is through assessment that teachers will provide feedback to their students and also make decisions about students' futures, it is important that teachers are also able to analyse the quality of these methods so they can rely on dependable evidence to make informed decisions.

3.2.4.1 Quality in classroom assessment

Knowing how to choose or develop different methods of assessment is not enough. Teachers must also be able to apply them properly and use them in ways that produce consistent results (AFT, NCME, & NEA, 1990). If classroom assessment information is of poor quality or incomplete, a teacher will not be able to effectively interpret and use information about students' learning (Brookhart, 1998). Therefore, for teachers to be able to produce informative feedback that is also relevant and according to students' needs, they will need good data about their students progress (Nicol & Macfarlane-Dick, 2006), which will depend on the quality of their assessments.

However, quality in classroom assessment needs a conceptualisation that goes beyond the traditional measurement theory (Black et al., 2011; Brookhart, 2003; Heritage, 2007; Mansell, James and the Assessment Reform Group, 2009; McMillan, 2000; Smith, 2003). First because classroom assessment is not only related to measurement, although measurement is an important component of assessment. Second because the dynamics of teaching and learning, and consequently of classroom assessment, do not allow teachers to go over all steps that are preached by the traditional measurement theory (analysis of internal consistency, pilot test, etc.).

That is, although classroom assessment should include the ideas of reliability and validity, it should be more closely related to an effective system of assessment, which produces dependable information to be used for instructional decisions, accountability and more importantly, to improve students' learning. However, both terms should be taken differently regarding formative and summative assessment (Black & Wiliam, 2012; Stobart, 2012). For summative purposes, it is the inferences that are most important, while for formative purposes, it is the actions that are most important (Wiliam & Black, 1996).

If FA is understood as assessment that is used during the teaching and learning process to explore students' understanding to help them develop that understanding and move forward in their learning, FA is valid if it fits-its-purpose, and therefore, if the purpose is to stimulate further learning (Stobart, 2012). Therefore, it will be reliable to the extent that it is being used to generate evidence that consistently leads to better, or better founded decisions (Black & Wiliam, 2012).

Shepard (2002) even argues that, in this case, reliability is less critical, since errors in instructional decision can be rectified through additional information gathered during another opportunity. However, if reliability does not matter for a specific assessment event, it surely still matters in the longer term. That is, what matters is not reliability of an individual decision, but the reliability of a longer term unit of analysis, which is the teacher's judgement over a number of assessment events.

On the other hand, when assessment is to be used with a summative purpose of grading, accountability, reporting or making placement decisions (Brookhart, 2011), reliability should be understood as "the extent to which an assessment can be trusted to give consistent information on a pupil's progress" whereas validity should be understood as "whether the assessment measures all that it might be felt important to measure" (Mansell, James and the Assessment Reform Group, 2009, p. 12).

However, in both cases:

> Quality assessment will contribute to improved learning – and it should do this through feedback and opportunities for the students to reflect on their performance, perhaps in consort with their peers. [...] Quality assessment will also assess all relevant types and outputs of learning, and not just the easy target of knowledge recall. [...] And quality assessment will do these things validly and as reliable as appropriate for its purpose and for the learning it addresses. Finally, quality assessment will be an 'open book' to everyone – its strengths and weaknesses in interpretation and dependability will be clear for all to see and, most importantly, to understand. (Gardner,

2012, p. 118)

Thus, to be able to provide effective feedback to their students, teachers need to have a deep understanding of these ideas. It will help them to choose or develop proper assessment methods and to analyse whether the assessment information gathered through them supports the intended purpose and use. In other words, to have a well-functioning assessment system, teachers need to see the process as a whole, which involves a kind of anticipatory thinking. When designing assessment tasks, teachers need to anticipate what their intentions are. They need to ask themselves questions like: What am I going to learn? What are students going to learn from this task? How am I going to mark it? How am I going to comment on it? Is the task well-designed to allow me to do that?

Therefore, the design of quality assessment involves a kind of anticipatory thinking about interpreting and communicating the data collected from that assessment, which I discuss next.

3.2.5 Interpreting and communicating assessment information

In this section, I present and discuss issues related to teachers' use of assessment information. Although I have discussed the importance of seeing the process as whole, and therefore, that the results teachers will get should be a reflection of a well-planned and well-designed system of assessment, the focus of this section is on how important it is for teachers to understand that they should be able to translate assessment data into usable information for different purposes and users (Brookhart, 2011; AFT, NCME, & NEA, 1990; Stiggins, 2011). As discussed previously, this information can come from formally gathered information, but also through informal observations and classroom discussions. This translation can be made through feedback to their students, but also, for example, through grades and reports.

In a way, they complement each other. The process of feedback will serve to provide teachers and students with the assessment information they need during the learning process, to make decisions that will help them with the improvement of learning. However, at some point, a summative judgement will have to be made (and classroom assessment should have room for it), and this is the information that parents and other lay audiences are usually interested in.

In class, based on what students say, for example, teachers can address misconceptions, lack of clarity or completeness, and level of understanding of different topics. Teachers should have the skills to interpret information while they walk around (D. J. Clarke, 1992) observing their students and during classroom discussions. These immediate interpretations are crucial for immediate actions. Teachers should know how to interpret and mediate the situation in a way that accepts students' answer, but also encouraging them to reflect that the solution presented is not consistent with the proposed task (Muniz, 2009). As pointed out by Ball et al. (2008, p. 401):

> During a classroom discussion, a teacher must decide when to pause for more clarification, when to use a student's remark to make a mathematical point, and when to ask a new question or pose a new task to further students' learning. Each of these decisions requires coordination between the mathematics at stake and the instructional options and purposes at play.

Furthermore, teachers also need to be aware of the questions 'not asked' by students, because they are also a very good source of evidence. Some students demonstrate that they are not learning, or do not see themselves as able to learn that content, by being quiet and not taking part in classroom discussions and/or activities. In this case, teachers must have the skills to detect the problem and encourage students' educational development and not to increase students' anxiety inappropriately (AFT, NCME, & NEA, 1990).

For parents, teachers should be able to explain in a family-friendly language their child's learning status in relation to age- or grade-level expectations (Stiggins, 2011), or the meaning of the mark that their child was allocated at the end of that term. In other words, teachers should have the skills to interpret and communicate that although the mark shows students' achievement during that period, it is also the reflection of what the child has already learnt and what s/he still needs to learn. That is to say, although teachers have to provide a mark at the end of a period, and that mark will vary from 0 to 100%, teachers must be able to explain that this mark is the result of a calculation and that in fact, the main objective was to evaluate various skills and competencies within a specific content.

Communicating to parents in Brazil can be a very hard task, especially for teachers from state schools, where most students come from the lower-income bracket and most parents did not have the opportunity to go to school. In this case, teachers should have the skills to communicate their educational decisions based on their interpretations of assessment results in a way that all parents will be able to understand and also take action on.

For coordinators, principals and other lay audiences, teachers can also be required to provide the proportion of students meeting or not meeting the expected target. Additionally, teachers can be required to explain whether the evidence was generated from different methods, or even whether the information was generated by comparing one student's work with that of other students (norm-reference), or with success criteria (criterion-reference), or with work of a similar nature previously produced by the same student (ipsative-reference) (Dudley & Swaffield, 2008).

Therefore, being able to translate assessment data into usable information for different purposes and users, in many cases, involves good written, but also oral and interpersonal communication skills (Brookhart, 1999). Regarding the latter, those who are involved should (Stiggins, 2011, p. 272):

- Understand what we are trying to accomplish by means of our communication about students' achievement.
- Use a common language to share information.
- Take time to tune in and be in the moment when information is being shared.
- Check back with each other in order to be sure everyone understood and felt able to use the achievement information shared.

Equally, teachers should also use the information from assessments to evaluate their own work, reviewing and restructuring it when necessary. Therefore, when getting inconsistent results, teachers should seek other explanations for the discrepancy or obtain other data to attempt to resolve the uncertainty, before arriving at a decision (AFT, NCME, & NEA, 1990): review the assessment method used, go through the learning intentions again and verify whether they were explained clearly and students understood what was expected from them, among other things. Thus, the evidence will serve to monitor students' learning, but also for teachers to reflect on their own strategies and to think about what they have provided to obtain that result. This will include both the method used to assess and how well they have taught the content to be able to conclude whether it was a problem with measurement or with learning.

In summary, teachers' use of assessment results should be done with a very clear intention and also make sense for those receiving the information. However, all the skills discussed in the previous sections will lose their purpose if teachers do not act in an ethical and legal manner.

3.2.6 Ethical and legal responsibilities

A final important aspect of FA is adherence to the highest possible standards of professional conduct (Morris, 1998), which involves the exercise of constant vigilance to ensure that every act performed when assessing formatively represents exemplary ethical behaviour.

For teachers, this 'exemplary ethical behaviour' relates to the whole feedback cycle, which includes planning the assessment tasks to collect evidence of students' learning (linked with clear learning intentions), interpreting this evidence, using the resulting assessment information according to the intended learning, and communicating the results to different audiences. More broadly, ethical behaviour also requires teachers to ensure that students have an opportunity to learn what is intended, and are suitably prepared to undertake the assessment tasks.

However, in the practice of assessment, be it for formative or any other purpose, this ideal can be challenged by a variety of political, personal, and financial factors that could compromise the high ethical standards that must be maintained if assessment is to best serve its purposes in learning.

For example, although 'teaching to the test' is a practice that is seen as unethical, especially if addressed to small proportion of students within a classroom, teachers can be making use of this practice because the expectations on them are sometimes so high in relation to summative assessment results that they see themselves as responsible for 'doing well', even though they are aware that the results will not reflect what students have actually learnt.

Therefore, when I refer to ethical and legal responsibilities, the purpose is not to establish rules on how to behave or solutions for every moral problem when assessing formatively, but to encourage teachers' reflection about general principles of moral behaviour and consequently how to act in problematic situations. These competences of ethical reflection are not specific parts of FA literacy, but professional qualities that should go alongside it in any teaching act.

Teachers should understand that proper ethical conduct can avoid damage in the educational relationship (with students, parents, authorities), as well as damage to education itself – avoidance of which should be the ultimate goal of the teaching activity.

Fairness is an important characteristic of ethics in assessment (McMillan, 2000). It includes avoiding stereotypes and taking into account students with disabilities and special needs, and also those with different social, cultural and economic backgrounds. "In essence, [fairness] indicates that each student must receive the same chance to demonstrate their knowledge, skills, abilities, and competencies" (Segers, Dochy, & Gijbels, 2010, p. 200).

Protecting confidentiality is another major responsibility of assessors (Brookhart, 2011; Morris, 2000; AFT, NCME, & NEA, 1990), and it is a responsibility that is often challenged in teachers' daily practice. An example of this is when teachers discuss an individual student's work in front of the whole class even if the intention is to help the group improve their learning. To avoid this practice from embarrassing students or violating their right of confidentiality, teachers can, for example, make a collection of students' mistakes and present them to the whole class without mentioning which student made the mistake.

Furthermore, following the general principle of ethics, in addition to their own practices, teachers should also attempt to discontinue any inappropriate practice of others, whenever possible (AFT, NCME, & NEA, 1990). However, it is important to note that although it is important that mathematics teachers observe such responsibilities, they do not really have a specifically mathematical aspect.

Finally, teachers should also know and observe the laws and guidelines which affect their city, state or district. In the next section, I present and discuss these documents in the light of what has been discussed so far.

3.3 Institutional literature

In this section, my aim is to explain to what extent the institutional literature recognises the recommendations from the research literature, which in turn will provide information in relation to what it is reasonable to expect from teachers.

I will discuss the documents that regulate and are guidelines for the country as a whole, but also those which refer only to the FD, which is the place where this study was conducted[4].

3.3.1 Brazilian level

In Brazil, there are two main documents that need to be analysed: the National Curriculum Guidelines for Basic Education[5] (SEB, MEC & DICEI, 2013) and the National Curriculum Parameters (PCN)[6].

3.3.1.1 The National Curriculum Guidelines for Basic Education

The Curriculum Guidelines is a 565-page document which includes ideas from the LDB-96 and specifies how states and municipalities should structure their education system. Only two pages refer to assessment[7]. In this context, although in a very brief and simplistic way, the emphasis in on FA, but not excluding summative practices.

It clearly states that assessment should "have a formative and participatory nature, be continuous, cumulative and diagnostic" [Art. 32 – I][8], and that it should "give priority to qualitative aspects of student learning above quantitative ones and also to results *during the process of teaching and learning* instead of emphasising results of final exams" [Art. 32 – III, my emphasis].

It highlights that assessment should not just be used to provide information about students' learning, but also "to detect problems in the process of teaching" [Art. 32 – Ia], showing the importance of teachers also using results of assessment for their own learning.

In terms of more specific elements, the document mentions that it is important to have clear learning intentions: "schools should express clearly what is expected of students in relation to their learning" [Art. 34]. However, it mentions that it is a role of the school and it does not specify how and to whom these intentions should be addressed.

Equally, it also recommends that teachers should "use varied instruments and procedures, such as observations, descriptive and reflective records, individual and group work, portfolios, exercises, tests, questionnaires, among others, with regard to its

[4]The reasons for choosing the FD are explained in Section 4.4.
[5]Translated from 'Diretrizes Curriculares Nacionais da Educação Básica' by the author.
[6]Translated from 'Parâmetros Curriculares Nacionais' by the author.
[7]Pages 137 and 138, Art. 32 – 35.
[8]All translations in this section were made by the author.

suitability to age and developmental characteristics of students" [Art. 32 – II], which has been largely discussed in the research literature.

However, it is important to notice that the recommendation is towards the "suitability to age and developmental characteristics of students", which can be explained by the fact that it is a common practice in Brazilian schools to mix all ability levels in the same class, including students with special needs.

This idea is also corroborated by another recommendation that says that assessment should "support teachers' decisions about their strategies and approaches according to the needs of students, create conditions to intervene immediately and also to remedy difficulties and redirect instruction" [Art. 32 – Ib].

At the same time, it views teachers as responsible for students' assessment when they mention: "the assessment of students is *to be held by teachers* and schools as part of the proposed curriculum, and as a guideline for pedagogical action" [Art. 32, my emphasis].

The system also recognises the importance of communicating assessment results to students' families when it mentions that assessment should also be used to "keep the family informed about students' performance" [Art. 32 – Ic] and "to recognise students and their families' rights to discuss results of assessment, and review procedures when their demands are well-founded" [Art. 32 – Id].

There are four other items that are specific to the Brazilian context which need explanation. They follow from Art 32, which states: "assessment of students is to be held by teachers and the schools as part of the proposed curriculum, and as a guideline for pedagogical action. It should:"

IV. Ensure different times and spaces for lower-achieving students to be supported properly throughout the school year;

V. Provide, compulsorily, opportunities of recovery, preferably throughout the school year;

VI. Ensure different times and spaces to teach students with poor attendance, to avoid, whenever possible, retention for absences;

VII. Allow progression of students who are not in the expected grade for their age.

Overall, these four items are very closely related. They all refer to actions to be taken in relation to students who are not progressing as expected. Item VII refers to a very problematic situation in the Brazilian system. According to the 2013 Census, approximately 25% of students are not in the expected grade for their age. At the same time, items IV, V and VI deal with the two most common reasons that put students in this situation (Moreira, 2013): absent students (item VI) and lower-achieving students (items IV and V).

In relation to item VI, the government has been taking some actions to help in overcoming the drawback of students' absences. One of them is the 'Bolsa Família Programme'[9], which provides financial aid to poor families. In return, families must ensure that their children attend school and are vaccinated. Research has shown that this programme has increased the attendance of students (Cardoso & Souza, 2004; Melo & Duarte, 2010), and also has had a positive impact on students' learning (Simões, 2012).

However, the recommendation in item VI is directed at those students that are still not attending school as expected. The recommendation is that teachers should ensure that these students are still taught the content that they missed when they were absent from school, so it is not necessary to make them repeat the year because of their absence.

Item IV is also related to opportunities that teachers and schools should give to students, but in this case, to the lower-achieving ones. One of these opportunities is the recovery mentioned in item V. As the item does not specify what it is meant by 'opportunities of recovery', most teachers interpret it as a second opportunity for students who did not achieve the expected mark to be able to recover the mark. Based on that, this is usually made through tests, at the end of each term. However, if at the end of the school year, the student has still not achieved the mark, s/he will have another chance to recover it through a final exam, which gives a summative connotation to this recovery.

Finally, the document specifies that "assessment procedures adopted by teachers and schools should be articulated with external assessments, which were created to help schools improve the quality in education and also students' learning" [Art. 33]. It also recommends:

I. Analysis of student performance based on results produced by external assessments should assist school systems and school community to rethink the educational practices to achieve better results.

II. External assessment of students' performance refers only to a restricted portion of what has been taught in schools. Therefore, teachers should still have the official

[9]Bolsa família can be literally translated as Family Allowance.

curriculum as reference, and not reduce their teaching to what is assessed by these large-scale tests.

Based on these two items, it is possible to infer that the system corroborates the literature towards the importance of administering and interpreting the results of external assessments (Brookhart, 2011; Stiggins, 1991; AFT, NCME, & NEA, 1990).

However, as explained in Section 2.4, the frequency with which external assessments are applied and their results disclosed, hampers teachers in using the evidence collected for improving students' learning. Here are the reasons:

1. In Brazil, external assessments are applied only in the last grade of each level, which means that most students will change schools at that point, not giving time for teachers to act for the benefit of students' learning.

2. It takes almost a year for the results to be disclosed.

3. The results are disclosed only in relation to the whole school, not for each student separately, which again does not enable teachers to act upon the results.

As the recommendation is to "assist school systems and school community to rethink the educational practices to *achieve better results*" [my emphasis], it can influence teaching and consequently the kinds of assessment that teachers use within the classroom, especially if the school has not achieved the expected results in the last year. In other words, teachers see themselves as preparing students for this kind of assessment.

Against this, the second part of the recommendation indicates that teachers should not teach to the test, because these tests refer to a much smaller part of the contents being taught in schools. However, the societal expectation is still towards good results in those exams. Therefore, in most cases, teachers still see their responsibility, within such a system and society expectations, as preparing their students for this kind of assessment.

3.3.1.2 The National Curriculum Parameters

The official guidance relating to mathematics, the PCN[10], was created by the government in 1998 (SEF, 1998). The aim was to provide elements for broadening the national

[10]Translated from 'Parâmetros Curriculares Nacionais' by the author.

debate about teaching and learning of mathematics, and also as an attempt to redirect the situation of mathematics education in Brazil, which could be defined as traditional and marked by an emphasis on the early formalisation of concepts, an excessive concern with skills training and the mechanisation of processes without comprehension.

For this purpose, it suggests problem solving as the starting point of mathematical activities, emphasising the importance of the history of mathematics, as well as the use of Information and Communication Technology (ICT) and games in mathematics teaching. Most of the recommendations in relation to assessment also follow this same idea:

> It is necessary to rethink some ideas that predominate in the meaning of assessment in mathematics nowadays, i.e. those ideas which conceive as priority only to assess if students memorised rules and schemes, not checking if they understood the concepts, the procedures that they used and the creativity in their solutions (p. 54).

The recommendation is that assessment should have a pedagogical, but also a social perspective. From the pedagogical perspective, assessment should provide teachers with information on how learning is taking place – the knowledge acquired, the reasoning been used, students' beliefs, habits and values, the appropriate use of strategies – so that they can review and redirect instruction of concepts and procedures only partially learnt.

From the social perspective, assessment should provide students with information about the skills and competencies that are socially required, whereas for teachers, it should help to identify which objectives have been achieved, to recognise the mathematical ability of students so that they can be inserted in the labour market and participate in the socio-cultural life.

In both cases, the recommendation is for the use of different methods (oral, written or visual representations) to provide teachers with information about students' skills to solve problems, to use mathematical language appropriately to communicate their ideas, and to develop and explain their reasoning.

These ideas should be linked with clear learning intentions – as indicators of the possible learning expectations which students are required to develop – in a way that:

> [...] the mathematical knowledge is seen as a meaningful construction, that learning intentions acknowledge the possible connections between concepts,

> and also the different ways in which these concepts can be applied; and that the analysis of a student progress takes as reference the student him/herself, and not only his/her position in relation to his/her peers (p. 55).

The document also highlights the importance of teachers using observations and dialogue to identify and interpret evidence of the skills developed by students (especially through students' mistakes), to judge whether the intentions previously outlined are being developed satisfactorily or if it is necessary to reorganise teaching to make this happen.

In summary, the document considers that teachers should change their view of how mathematics is learnt and taught and consequently assessed. However, the ideas are discussed in a disconnected way and there is no indication of how this should be done. Equally, there are important aspects that they do not mention. The students' involvement in the assessment process, for example, is one of them.

In the next section, I present and discuss the official document from the FD.

3.3.2 Federal District level

There is one document in the FD which specifies how assessment should be conducted: the Guidelines of Educational Assessment (GEA) (GDF, SEEDF & SUBEB, 2014). The document clearly explains which conception of assessment will underpin the ideas that will be proposed:

> Assessment is not limited to the application of tests and exams, and cannot be interpreted only as measurement. Measurement is just a part of assessment. The most important part, actually, is to analyse the information gathered through assessment to make constant interventions [...] assessment must be seen as a central issue in the process of teaching and learning, and must be useful for teachers and students [p. 8].

Based on that, the GEA specifies that FA is the form that relates the most to this conception of assessment. The GEA then provides the concepts and practices that underpin this conception. In other words, the salient point of the GEA (as opposed to the other two documents presented above) is that it not only proposes changes in the way assessment has been conducted so far, but it also provides suggestions as to how it could be done.

Alongside these, the difference between formative and summative assessment is stated.

> This is the meaning of *assessment for learning* as opposed to *assessment of learning*. The difference is that the former should be used to promote interventions *while* the process of teaching and learning occurs and the second, also called summative assessment, sums up the learning that occurred *after* a certain period of time and is not always used for interventions [p. 9].

In addition, this definition is expanded by pointing out that there are some elements that are essential for teachers to be able to put FA into practice: feedback, self- and peer-assessment, the use of different methods and involving students in this process.

In relation to feedback, the GEA explains that it is "indispensable for FA to occur", because feedback enables "those who are involved to understand what they have learnt, what their weaknesses are and also helps them to develop self-regulation skills" [p. 11].

The GEA does not explain how feedback should be provided and the kind of information that teachers should include in it, but it mentions that feedback is a form of "constant dialogue between teachers and students" [p. 15] and that it should "be provided as soon as possible so students are able to act upon it and teachers are able to plan further interventions" [p. 31]. Additionally, it specifies that when teachers are giving feedback to students, they must also point out students' strengths, and not only their weaknesses.

The use of self- and peer-assessment is also encouraged throughout the GEA. As well as feedback, self-assessment is seen as essential in FA, and it is defined as:

> [...] the process in which the student him/herself continuously analyses the activities developed and those still being developed, records his/her perceptions and feelings and identifies future actions needed. This analysis takes into account: what s/he has learnt, what s/he has not yet learnt, the factors that were facilitating or hindering their progress – having as reference the learning intentions and the assessment criteria [p. 33].

It is also recommended that teachers should not encourage students to give marks for themselves, which has also been shown to be a common practice in Brazil (Villas Boas, 2008). The GEA explains that by giving marks, students will not pay attention to what they missed and what still needs to be done, and therefore will not be using self-assessment for the improvement of their learning.

Peer-assessment is seen from the same perspective and also as a good way of developing self-assessment skills in students. As suggested by the research literature (Black et al., 2003; Stiggins, 2011), the GEA highlights the importance of teachers encouraging self- and peer-assessment constantly because “when students assess themselves and assess the work of peers, it contributes to their intellectual and personal growth, at the same time as it empowers their learning in a collaborative and purposeful way” [p. 18].

Regarding assessment methods, the GEA emphasises the importance of using different forms to assess different skills required from students. Additionally, the GEA makes it clear that “it is not the method that will guarantee that FA is happening, but the intention that the assessor gives to them and also how they will use their results” [p. 9]. These methods must also “reflect the learning intentions and the success criteria that were preferably negotiated with students” [p. 12].

Again, it is also emphasised that different methods should not only be used for the attribution of marks, but to collect evidence of what students have learnt and what they still need to learn, and which kind of intervention will be needed.

As tests and homework assignments are still the methods most commonly used for assessing students, the GEA has specific sections for these two practices.

In relation to tests, it is first specified that they are not against the use of them when assessing students’ learning, but it is suggested that if teachers decide to use tests, they should “be articulated with other methods, must take learning intentions into account, and should be linked with immediate feedback to students” [p. 31]. Therefore, even when teachers use tests (which is usually linked to a summative assessment), they must also be used with formative purposes.

It is also acknowledged that, based on how the Brazilian system is organised (in which teachers are required to give marks so the school can decide whether or not a student will progress to the next grade), it is expected that teachers will place a lot of emphasis on marks. However, the GEA suggests that “these marks should not come only from tests results, and when tests are used, it should not count as more than 50% of the mark”. It is also emphasised that the marks that are given throughout the year should be used as “additional evidence of students’ learning” [p. 32].

However, the GEA reiterates that “the mark itself does not provide sufficient information” and therefore “teachers should be able to provide evidence in relation to the different actions that were taken to come up with this numerical symbol" [p. 32].

Similarly, when referring to homework assignments, the GEA emphasises that it is necessary to move from the current practice – assigning as homework only the exercises from the textbook that there was not enough time to complete during the lesson – to assigning "significant activities, based on clear learning intentions, in reasonable amounts and according to the level of each student" [p. 22].

Furthermore, the GEA argues that teachers need to change their actions after students complete the homework assignment. The current practice is to register whether students have done or not the assignment (with those who have not done it usually penalised in their final marks) and then for teachers to 'correct' the exercises on the board for the whole class. The suggestion in the GEA is to move from this teacher-centred approach to a student-centred one, in which peer-assessment can be encouraged or students are invited to show their solutions to whole class, explaining their reasoning and steps followed to arrive to the final solution.

The 'recovery', which was mentioned in Section 2.4, is also discussed in the GEA, specifying how it should be done: "it is recommended to carry out continuous interventions for all students, as soon as their learning needs were evidenced and met [...] self-assessment has an important role in this process" [p. 23].

It is highlighted that teachers need to change their view from the idea that the purpose of the 'recovery' is to give students another opportunity to recover their mark, to the view that it is about improving learning. Therefore, it is recommended that "if teachers use the evidence from assessment continuously, they will be able to remedy students' problems while learning is happening, which in most cases will make it unnecessary to recover the mark at the end of a period" [p. 24].

The GEA also raises three other very important issues, which largely influence assessments (and consequently teaching and learning) in the FD: the 'coordenação pedagógica', the 'conselho de classe' and the 'regime de progressão parcial com dependência', which were not mentioned in the other documents because they are not compulsory for every state in the country.

As I explained in Section 2.3, the conditions of teaching in the FD are much better than in the rest of the country. The teacher has a workload of 40 hours, divided between lessons, and meeting and planning hours. These meeting and planning hours are what we call 'coordenação pedagógica'.

There is no rule as to how this 'coordenação pedagógica' should be structured. However, most schools allocate one day per week to meet parents (when parents request this or when teachers think it is necessary) and another day for a meeting with the whole group of teachers. On the other days, teachers are free to decide what they will do. A small number of teachers use this time to meet with lower-achieving students to help them with their needs.

In the GEA, the suggestion is that the ‘coordenação pedagógica’ should be used for teachers “to discuss their conceptions and practices of assessment, and also as a good opportunity for teachers’ self-assessment” [p. 8], among other things. Teachers should use it “to help each other, especially in relation to developing assessment methods with clear learning intentions in which the results will be used to provide feedback that moves students’ learning forward” [p. 31].

The ‘conselho de classe’ is another meeting that assembles teachers and school management teams together. It is held at the end of each term and the current practice is to analyse and list which students have not achieved the average mark, i.e., it is used only for accountability purposes. Based on this analysis, the school usually contacts students and their families and explains the results of the ‘conselho de classe’, placing on the families the responsibility of doing something so the student can achieve the mark in the next term. The ‘conselho de classe’ held at the end of the school year is also used to decide which students will progress to the next grade, those who will have to repeat the same grade next year and those who will be included in the ‘regime de progressão parcial com dependência’.

However, the recommendation is to use the ‘conselho de classe’ to “identify what students have learnt, what they have not learnt and what must be done for them to learn what they missed. It should involve students’ families, all the school staff and students themselves.” This should be done in such a way that it does not “expose, label, punish and exclude students” [p. 28], but to help them in their weaknesses.

Equally, the GEA explains that the ‘regime de progressão parcial com dependência’ should follow the ideas explained in Section 2.4.1, but that teachers should also take this opportunity to put the ideas of FA into practice.

> [...] for it not to become pseudo-approval, or worse, pseudo-learning, the methods used must be carefully designed to meet students’ needs. The assignments must be linked with clear learning intentions and must be designed to improve students’ learning so that they will not face the same problems in the current grade [p. 25].

Finally, it is also emphasised that assessment needs to be conducted ethically, which means taking into account the following:

- respecting students' work/productions;
- not using assessment for comparison (students' progress should be compared with their own abilities and not with those of peers);
- using assessments to encourage students' progress and not to threaten, embarrass and punish them;
- using assessment results for the improvement of learning (without being included in any form of ranking), and
- disclosing assessment results only to students and their parents/guardians [p. 34].

Overall, the GEA argues that assessment should follow a formative approach, in which students' learning must always be the ultimate goal, which is in accordance with what is suggested by the research literature.

3.4 Summary

In this chapter the focus was to discuss the knowledge and skills that it is desirable for teachers to have for successful implementation of FA practices. The pieces of research summarised in this chapter show that there are several aspects that need to be addressed when referring to FA.

Similarly, the analysis of the Brazilian and FD official documents has also shown that they are in agreement with the importance of these ideas and recognise that changes are still needed to put these ideas into practice.

In the next chapter, I detail issues and questions that emerged from the literature review followed by an overview of how I designed this study to address them.

CHAPTER 4

RESEARCH DESIGN OVERVIEW

The previous chapter examined elements of FA identified in the research literature, and compared to the institutional literature. These elements are numerous, and a careful selection process must occur to be able to design further investigations. This chapter outlines my decisions, with respect to the research design, to address these elements.

Section 4.1 introduces and explains the research questions which influenced the choice of the research paradigms presented in Section 4.2. Section 4.3 provides some epistemological considerations, followed by Section 4.4 which explains the choice of location and participants. Section 4.5 presents an account of the ethical issues involved in this study.

4.1 Research questions

Assessment impacts on millions of children and educational professionals, and in Brazil it is no different. To progress as a society we need a quality education system, which depends, among other things, on a professional body that is able to recognise the importance of assessment and its impact on student learning.

There are several elements of assessment that can be included with reference to teachers' knowledge and skills. Researchers and those responsible for teacher training have devoted a wealth of writing to the impact of assessment on students and teachers, and made multiple formal calls for research in this area.

Although all the ideas presented in the previous chapter are equally important, given the time scale available for a PhD project, and to increase the feasibility and depth of the study, I decided to focus on a smaller number of critical elements, to answer the following question:

> What do Brazilian secondary mathematics teachers know about FA in general and the idea of feedback in particular?

It is widely accepted that what teachers know is one of the most important influences on what is happening in the classroom and that they "need substantial knowledge to implement FA effectively in classrooms" (Bennett, 2011, p. 20). Regarding this, and based on my experience with Brazilian teachers outlined in the introduction of this dissertation, I share Shepard's (2002) view that teachers' knowledge should be a primary, although not the only, site for research.

While there are different conceptions of teachers' knowledge, my intention was to understand the 'knowledge-for-practice' (Cochran-Smith & Lytle, 1999) of new and experienced mathematics teachers in terms of what, generally speaking, is already 'known' about such knowledge based on research studies. Therefore, in this study, I did not challenge this 'research-based knowledge', but accepted the recommendations that have been made by the research literature and recognised by official documents (GDF, SEEDF & SUBEB, 2014; SEB, MEC & DICEI, 2013; SEF, 1998), and used these to develop a framework that specifies domains of mathematics teachers' FA knowledge.

Based on significant evidence that much of this research-based knowledge is unfamiliar to teachers (Black & Wiliam, 1998; Campbell, 2013; Plake & Impara, 1997; Stiggins, 1991, 2002), I anticipated, that this would also be the case with mathematics teachers in Brazil, and that I would find confirmation of some kind of gap between what they know and what, according to the literature, they 'should' know. As the literature has already shown that teacher knowledge-for-practice does not (generally) take (much) account of research-based knowledge, I wanted to understand the nature of this gap.

My intention was not to criticise individual mathematics teachers by pointing out what they know or what they do not know, but to understand their knowledge of FA. The desired outcome was to be able to feed back to the teachers themselves, and those working with them, about where and what the gaps are to close them.

Therefore, I aimed to answer the following questions:

> Are there any systematic patterns of variation in teachers' knowledge:
>
> 1. relating to different aspects of FA?

2. according to background characteristics of the teacher (e.g. years of teaching experience or level of education)?

In aiming to answer these questions, I sought to obtain a more nuanced account of teachers' FA knowledge. However, the way that assessment has been discussed is decontextualised from what it is to be a mathematics teacher. Moreover, no relevant instrument was available which was suitable for Brazilian mathematics teachers. Therefore, I had an additional goal of analysing:

> To what extent is it possible to produce an instrument which reflects in a practical way what has been said in the literature of FA?

In aiming to answer this third research question, I wanted to analyse to what extent it was possible to produce a structured instrument to assess mathematics teachers' knolwedge of FA.

Therefore, in the next sections and chapters, I will not only discuss the results in relation to their knowledge, but also what it meant to take the recommendations made by the literature and turn them into the Mathematics Teachers' Formative Assessment Knowledge instrument (MaTFAKi). I start by explaining the research design.

4.2 Research design

The literature suggests that the choice of the research design and methods should start with the research questions and the aims of the study (Cohen, Manion, & Morrison, 2011; Oppenheim, 1992; Robson, 2011).

As my main research question aimed to investigate 'what' Brazilian mathematics teachers know about FA, it was of a descriptive nature and therefore, survey research seemed to be appropriate (Bell, 2014; Cohen et al., 2011). My main intention with this study was to 'survey' Brazilian mathematics teachers to provide a description of their knowledge of FA. Survey research fitted my purpose in testing the hypothesis (Creswell, 2014) that there was some kind of gap between what teachers know and what they 'should' know. In addition, it enabled the generation of standardised information in the form of numerical data (Morrison, 1993), which was not available about mathematics teachers in Brazil.

In terms of methods, analysing studies that had 'assessment' as focus, the options varied from *(a)* observations (Lyon, Miller, & Wylie, 2011), *(b)* artefact-based measures

(Randel et al., 2011), *(c)* self-reports (which include some types of questionnaires, logs, and portfolios) (Kershaw, 1993; McMillan, 2001; Zhang & Burry-Stock, 1994), *(d)* interviews (Graham, 2005), and *(e)* tests (Mertler & Campbell, 2005; Plake, 1993), all of them with strengths and weaknesses.

With all those possibilities, and considering that no one type of instrument is universally preferable to another, what I needed to analyse was what type of instrument best suited the intentions of this study. According to Randel and Clark (2013), if the intention is to measure teachers' knowledge and/or reasoning across a broad array of assessment practices, a test might be the best option. However, if I also wanted to correlate knowledge and skills with actual classroom practices, I would have needed to observe some specific lessons. On the other hand, when the intention is to collect information from or about people to describe, compare or explain their knowledge, attitudes and behaviour, questionnaires tend to be the ideal method (Fink, 2008), which felt more aligned with the purpose of this research and the research questions that I set out to answer.

Therefore, I decided to use an online structured questionnaire with a quantitative approach, to collect a large volume of standardised data at a low cost, from a broad sample in a short period of time (Kelley, Clark, Brown, & Sitzia, 2003; Cohen et al., 2011). However, it is notable that the development of the MaTFAKi took me much more time than I planned, as I detail in Chapter 5.

4.2.1 An *online* questionnaire survey

As with any methodology, survey research has its drawbacks. In the same way that there were concerns about the appropriateness of mail and telephone surveys back in the 1970s, nowadays, web-based surveys raise further concerns (Schonlau, Fricker, & Elliott, 2002).

However, as the internet has inevitably become part of our lives, many guidelines have been written (e.g. Arsham, 2014; British Psychological Society, 2013) and studies conducted (e.g. Wright, 2005; Van Selm & Jankowski, 2006) to take these concerns into consideration.

In this study, one of the main reasons that made me choose an online questionnaire was the experience I had with Brazilian teachers on my MPhil and the first year of my PhD. In the MPhil, I had a great response rate when I asked teachers to answer an online questionnaire (35% as opposed to the usual 10–15%). In the first year of the PhD, I had disappointing participation when I tried a closer approach through participatory research. Additionally, because I was living in England, being able to administer the MaTFAKi online facilitated data collection.

Another important reason was in relation to the features that web survey environments can provide. Using a web-based environment allowed me to hide and show some parts of the MaTFAKi, which would not be possible in a paper-based questionnaire. The particular characteristics of the questionnaire developed for this study are explained in Section 5.3.

Access problems, which are often cited as an issue of web-based questionnaires, were not evident in this study as the target population had internet access in all schools. Computer expertise was not a problem as the MaTFAKi only required them to click on a link to open the MaTFAKi and choose a response option.

In the next section, I provide some epistemological considerations underpinning this study and how they influenced the planning and implementation of the research design and analysis of its results.

4.3 Epistemological considerations

Epistemology represents "the claims or assumptions about the possible ways of gaining knowledge of social reality, whatever it is understood to be: that is, claims about how what exists may be known" (Blaikie, 2000, p. 8).

In the MaTFAKi, the data collected was standardised and limited to the four options included in the questions. The main assumption underpinning my choice of method was that by taking appropriate steps to ensure the validity and reliability of the MaTFAKi (see Sections 5.2 and 6.3), and by drawing reflexively on my contextual knowledge as an insider in Brazilian schools, such an instrument could produce worthwhile results, capable of being understood and interpreted.

Nevertheless, these results might also be a reflection of what the questions were asking and how they were asked in the MaTFAKi. The MaTFAKi was developed based on theories and other studies, and was strongly influenced by my experiences as a researcher and mathematics teacher. Therefore, it was immersed in my interpretation of this literature and my own understanding of FA supported by my experience and knowledge of the Brazilian educational system. Thus, even though I was very careful in developing the MaTFAKi, respondents might have seen different meanings in the terms employed in questions and the response options. A way to address this concern would be to conduct several pilot tests, in the course of which this very issue would be subject to investigation.

The same can be said in regards to the construct being assessed by the MaTFAKi, as it was designed to measure teachers' knowledge of FA, it can raise a common dilemma about teachers' professional knowledge – that there might be a distance between knowledge that comes from research and knowledge that comes from teachers' practices (Loughran, 2012).

On one hand there is a danger that a research-based questionnaire about FA poses questions which are too decontextualised for them to be effective in eliciting teachers' professional knowledge. This was the reason for translating more theoretical knowledge into classroom situations resembling those that teachers were actually likely to experience.

On the other hand, presenting teachers with classroom scenarios posed the risk that teachers would answer according to what they would do, or are accustomed to doing in their own lessons (therefore based on their experience), instead of considering the (theoretical research-based) ideal attitude to be taken in that specific situation regardless of their working conditions or existing practice, even though I acknowledge the importance of teachers' specific context when doing educational research. This issue is discussed further in Section 6.2 where I detail the face-to-face review conducted with teachers.

However, as Shulman (1999, p. 62) indicates "teachers themselves have difficulty in articulating what they know and how they know it". By providing teachers with four options to choose from, the MaTFAKi represents one possible way of supporting teachers in showing what they know about FA. In this case, teachers' answers to these structured items help to make their professional knowledge more explicit and begin to unfold what the knowledge base of FA in mathematics teaching might look like.

Regardless, what was carefully considered was whether or not the methodology and method chosen would answer the research questions and would satisfy the aims and objectives of the research. Moreover, it needed to be taken into account that the choice of the methodology and methods inevitably leads to the generation of particular kinds of data and forms of analysis.

In the next section, I explain the choice and location of the participants.

4.4 Location and participants

This study targeted Brazilian mathematics teachers in secondary state schools. The choice of state schools was due to the fact that around 80% of Brazilian children and teenagers are in state schools, and because I taught in independent and state schools in Brazil and know how differently they operate, especially referring to how flexible and creative teachers' can be in their lessons. Private schools tend to be more rigid in that sense.

I focused on teachers in the FD of Brazil for three reasons. Firstly, Brazil is a huge country; while in the earlier phases of the study (see Section 6.3.1.1) securing a representative return was not essential, it was in the final phase. Thus, I decided to limit the scope of this final phase to make it feasible within a PhD project. Secondly, because I have been working in the FD since finishing my undergraduate studies, I understand how the system works. This contributed to the development of the MaTFAKi, enabled my interpretation of the data collected, and facilitated my access to teachers in the final phase. Thirdly, because of what is proposed in the GEA and the teaching conditions in the FD (see Section 2.3), I judged that I would be more likely to find teachers who could demonstrate at least part of the knowledge that is proposed in the literature.

Once the design and location of the research was defined, some important ethical issues were taken into consideration.

4.5 Ethical considerations

Considering that "Educational research undertaken by UK researchers outside of the UK must adhere to the same ethical standards as research in the UK" (BERA, 2011, p. 5) and the non-availability of such guidelines in Brazil when this study was planned and conducted, throughout the data collection, analysis, and reporting, I followed the ethical guidelines of the British Educational Research Association (BERA, 2011):

Voluntary Informed Consent: During the pilot phases, the information on informed consent was provided directly to participants. For the main survey, informed consent was collected both from the Department of Education in the FD and the teachers. The first action, therefore, was to contact the Department of Education to ask their permission to send the email to the schools. After receiving their consent, I contacted the schools by email, following the sampling and delivery strategy explained in Section 8.1. I introduced myself as a FD mathematics teacher and provided them with an explanation of the research and the importance of their participation. At the same time, I specifically stated that if they submitted their answers, they would be automatically agreeing to participate in the research and their answers would be included in the data analysis. My contact details were included and all e-mails received were answered kindly, giving explanations when required. Cover letters and written consent forms were provided in all phases.

Openness and Disclosure: In every phase, participants were informed about the purpose of the research and the importance of their participation. It also provided assurance that their anonymity and confidentiality would be respected, any data collected would be used for research only, and the results would be part of this dissertation and disseminated in academic events and research articles. The information provided in relation to the research was enough for the participants to understand what they were required to do, but not so much that it could influence their answers. For example, they were informed that the research was in relation to FA but were not told that it was in relation to 'their knowledge' of FA. Instructions on how to go about answering the questions were also provided.

Right to Withdraw: During the pilot phases, participants were informed about their right to withdraw or to participate only in parts of the research. They also had the right to not be recorded. During the main survey, teachers were informed that if they did not want to participate, they could either not start the questionnaire or give up in the middle. Their answers would not be recorded, unless they clicked on the appropriate button at the end of the questionnaire. Unfortunately, because the data was anonymous, it was not possible to exclude any answers once they were submitted. However, this was clearly explained in the cover letter.

Incentives: Incentives, using a lottery scheme, were offered to encourage teachers' participation (see Section 6.3.1). For this reason, I informed them about the process of prize drawing and the selection of eligible participants. To draw prizes I collected contact information from participants. This personal data was stored securely. As they were also invited to participate in further research, it was no longer possible to anonymise the responses for those who included their personal data. However, they were informed of this in the cover letter and again in the section of the MaTFAKi where they were supposed to provide this information.

Privacy: In all phases, I guaranteed the anonymity of every participant by assigning pseudonyms and did not reveal the names of their schools in any circumstances.

Finally, considering the **Responsibilities to Educational Professionals, Policy Makers and the General Public** (BERA, 2011, p. 10), it is my purpose (see Section 4.1) to make the results with this research public for the benefit of education professionals and policy makers. The findings will be communicated in an appropriate language and accompanied with suggestions of how the findings can be used by teachers and for the improvement of the provision of CPD courses.

4.6 Summary

In this chapter, I have provided an overview of the study, explaining what I wanted to achieve with my research questions and the decisions I made based on them.

I explained the choice of an online questionnaire as the main method of data collection, and outlined the epistemological and ethical considerations that were taken into account during all phases of this study with mathematics teachers in the FD of Brazil.

Following this overview, in the next chapters I provide a detailed account of the realisation of this research design.

CHAPTER 5

DEVELOPING THE MATFAKI

The previous chapter explained that one of the aims of the study led to the development of the MaTFAKi. This chapter explains how I structured each step taken before writing the items in the MaTFAKi.

Section 5.1 explains how six Domains of Knowledge (DoK) were chosen to be included in the MaTFAKi as being essential aspects of FA. Section 5.2 introduces the phases conducted to ensure the validity of the MaTFAKi in assessing the six DoK. Section 5.3 outlines the decisions made in terms of type of questions and the layout of the MaTFAKi. Section 5.4 explains the approach taken to write the questions and introduces one of the scenarios of the MaTFAKi to facilitate comprehension at this stage.

5.1 MaTFAKi conceptualisation

According to Radhakrishna (2007) and Rust & Golombok (2009), 'conceptualisation' is when content from the literature or theoretical framework is transformed into statements or questions. It was at this stage that I had to decide which aspects of FA were essential, and therefore, could be considered as a starting point to be included in the MaTFAKi and form the basis to develop the questions.

There are many lists in which authors and institutions summarise the knowledge and skills teachers should have to be able to assess well (e.g. Brookhart, 2011; Popham, 2009; AFT, NCME, & NEA, 1990; Xu & Brown, 2016). Based not only on these lists, but on a broader review of the research and institutional literature, I identified six DoK which would provide a good start for the development of the MaTFAKi. These DoK were later validated by a team of researchers from Brazil (see Section 5.2.1). In the next section I explain the thinking behind the choice and definition of each domain.

5.1.1 Defining the domains of knowledge

In this section, I explain the rationale for defining the six DoK that were used as the basis for developing the MaTFAKi. In the first part I introduce how I came up with each of them. In the second part, I present and explain each DoK separately.

In the framework specified in Section 3.2, I argued that feedback should be taken as the heart of FA, and not only considered as a component of FA.

Therefore, following this rationale, the first DoK emerged from the main idea of the framework. The 'Knowledge of providing effective feedback (KEF)' is essential for teachers to be able to consider feedback as the heart of FA. Following from this, and considering the three essential questions that 'giving feedback' encompasses (Hattie & Timperley, 2007), the other DoK arose from them:

Where am I going? When giving feedback, the teacher should be able to make clear to students what they are supposed to achieve. This idea generated the second DoK: the 'Knowledge of articulating and sharing clear learning intentions and success criteria (KLI)'.

How am I going? When giving feedback, teachers also should be able to make clear to students what they have already achieved. Therefore, teachers need to be able to elicit evidence of students' learning – the third DoK: 'Knowledge of designing and/or choosing assessment methods or classroom activities and discussions to collect evidence of students' learning (KAM)'. However, collecting evidence is not enough. To be able to answer the question, teachers should also be able to interpret the information elicited – the fourth DoK: 'Knowledge of interpreting evidence of students' learning (KIE)'.

Where to next? Finally, giving feedback involves more than just making clear to students what is expected from them (Where am I going?) and what they have achieved (How am I going?). In fact, assessment can only be considered formative if the evidence elicited is used by teachers, learners or their peers to make decisions about the next steps in teaching and learning (ARG, 1999, 2002; Wiliam, 2011). This involves the fifth DoK: ‘Knowledge of closing the feedback loop (KCL)’. Giving students the opportunity to act upon feedback means they are able to check whether learning occurred as a result of the feedback (Sadler, 2010). However, ‘acting upon the feedback’, among other things, does not come naturally to students. Therefore, the ‘Knowledge of helping students use assessment information (KHS)’ – the sixth DoK, is also essential for teachers when helping their students to understand what is missing to achieve the learning targets that were previously made clear to them.

Figure 5.1 tentatively illustrates the six DoK and their relationships explained above.

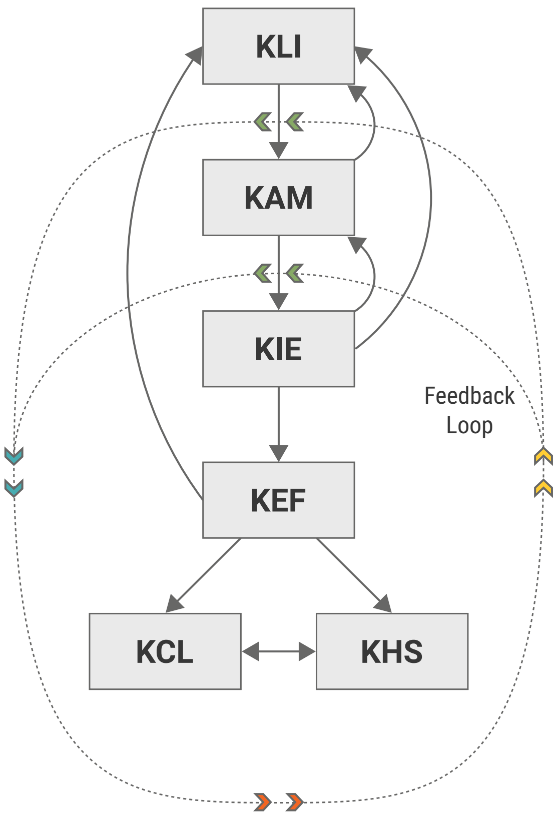


Figure 5.1: The Domains of Knowledge and their relationships.

As I explained, these domains were defined as the framework to develop the MaTFAKi; they are not supposed to represent all aspects that teachers should know to be considered FA literate. In the next section, I better explain what is included in each DoK and how they are related to teaching and learning mathematics. Although I present and explain them separately, I anticipate that these aspects are not independent, but inter-related and represent some essential aspect of the overall concept of FA.

In Chapter 7, where I detail the MaTFAKi itself, I better explain my rationale for transforming each of them into questions.

5.1.1.1 Knowledge of providing effective feedback (KEF)

As discussed in Section 3.2.1, considering feedback as the heart of assessment is not to reduce it to the mere provision of feedback. To be effective, it needs to move learning forward (Wiliam, 2011).

Research has shown that, to be effective, teacher feedback on student work should be descriptive and should comment on the work itself and the process used to do the work (Hattie & Timperley, 2007; Kluger & DeNisi, 1996, 1998). It should give students information about their work set against the criteria for good work, as articulated as part of their learning intentions and shared with students (KLI). Effective feedback is elaborated, but not too complex; is specific to the work; avoids general praise; and is different for different learners (Shute, 2008; Stiggins, 2009). Effective feedback is at an intermediate level of generality (Brookhart, 2008; Kluger & DeNisi, 1996) specific enough that students can identify the improvements needed, but not so specific that the work is already done for them.

In mathematics, Brookhart (2008) suggests that feedback could have, for example, the problem-solving process as a reference, and therefore should point out whether or not the student:

- identified the problem (What exactly is being asked here?)
- identified the elements of the problem (e.g. Which numbers need to be added? Which numbers are irrelevant?)
- identified one or more strategies that would work to solve the problem and then choose and apply one of these strategies
- evaluated the solution they arrived at to see if it is a reasonable response to the problem

The KEF – taking the problem-solving process as reference or not – is related to all the other domains, as I previously explained. In addition, being able to give effective feedback depends on teachers' subject knowledge, not only to make sense of what students say or have written (KIE) but also to be able to determine what would be the most appropriate next steps (Hodgen & Wiliam, 2006). For Ma (1999), this involves profound understanding of fundamental mathematics rather than the abstract knowledge of advanced study in mathematics. For Ball et al. (2008, p. 395), this is mathematical knowledge for teaching, or “mathematical knowledge needed to carry out the work of teaching mathematics”.

5.1.1.2 Knowledge of articulating and sharing clear learning intentions and success criteria (KLI)

One of the first principles of FA is clarifying and sharing learning intentions and success criteria (Black & Wiliam, 1998; Wiliam, 2011).

Therefore, teachers should be able to articulate what is expected for students to learn in clear, attainable and assessable ways (Brookhart, 2011). Clear, so students can understand what is expected from them. Attainable, so students can achieve them. Assessable, so that both teachers and students will know whether and to what extent they have been achieved. Therefore, they are crucial to answer the question ‘Where am I going?’ in the feedback cycle (KEF).

This should include starting from where the learner is. According to Hodgen and Wiliam (2006), this also helps overcome the idea of mathematics as being something disconnected and inconsistent because it will show students that they are reconstructing ideas they already have and not just overlaying them with new ones (like they were irrelevant in first place).

However, as argued in Section 3.2.3, being able to articulate and share the learning intentions and success criteria is not enough. Teachers should also be able to check how the students understand the targets, which involves how students interpret what it is meant by ‘good work’.

To be able to help their students move forward in their learning and to understand what makes a good piece of mathematics, teachers should provide much more than lists of criteria. They need to provide students with opportunities to engage in mathematical argument and reasoning so they and their peers can learn the ways in which the quality of mathematical work is judged (Hodgen & Wiliam, 2006). This idea is directly linked to the kinds of activities that teachers are using in their lessons (KAM) to elicit evidence, to understand where their students stand in their learning (KIE), and to identify the necessary next steps.

5.1.1.3 Knowledge of designing and/or choosing assessment methods or classroom activities and discussions to collect evidence of students' learning (KAM)

To be able to provide effective feedback (KEF), teachers need to know where their students stand in their learning, and for that, they need to be able to design and/or choose assessment methods or procedures to be able to obtain that information. Pellegrino (2006) contends that providing students with multiple opportunities to apply what they have learned is a superior approach to relying upon a singular assignment for the purposes of achieving that goal. For that, teachers should understand the purposes and uses of the range of assessment options, being aware that different methods/procedures can be incompatible with certain goals (KLI) and may impact quite differently on their teaching (AFT, NCME, & NEA, 1990).

For FA, teachers should design or choose methods/procedures from which both teachers and students can use the results (Black & Wiliam, 1998; McMillan, 2000; Stiggins, 2009). According to Brookhart (2011), a method constructed/chosen to be used formatively will have the main purpose of giving feedback to students (KEF) and will be based on a narrow range of learning intentions (KLI).

Therefore, these methods, which also include classroom activities and discussions, should enable teachers and students to understand what is needed next. Consequently, they should provide opportunities for mathematical learning which includes students being able to share their ideas (Hodgen & Marshall, 2005).

According to Hodgen and Wiliam (2006, p. 6), in mathematics it should include, among other things:

- challenging activities that promote thinking and discussion;
- encouraging pupil talk through questioning and listening;
- strategies to support all learners to engage in discussion;
- peer discussion between students; and
- rich and open whole-class discussions.

These activities will allow not only teachers, but students themselves to explore the mathematics in such a way that they will be better able to understand what they know, how well they know it (How am I going?) and in which areas more work is needed

(Where to next?). Using summative tests with formative purposes is another example of how teachers can use assessment methods to elicit evidence of students' learning and provide effective feedback based on that information.

5.1.1.4 Knowledge of interpreting the evidence of students' learning (KIE)

Knowing how to draw inferences from students' responses (oral or written) or knowing what students know (Pellegrino, Chudowsky, & Glaser, 2001) is crucial to the effectiveness of FA and consequently to giving feedback to students (Black & Wiliam, 1998; Hattie & Timperley, 2007).

Essentially, teachers should be able to identify students' current mathematical understanding so that they can modify instruction to facilitate improvement. Analysis of students' responses can be either on formal (usually written) assessment, which will usually take place after the lesson and therefore the teacher will have more time to think and closely examine it; or informal (usually through observations or classroom discussions), in which the teacher will have to make inferences (and interventions) on a moment-by-moment basis (Suskie, 2009).

The importance of engineering mathematical activities that provide actionable information to teachers was already discussed in the previous section (KAM). To be able to interpret the evidence generated through these activities, the teacher needs to be able to accept a range of responses (orally or written), be they right or wrong, and more importantly, take them all seriously so students are not discouraged to make mistakes and to help students to recognise inconsistencies, respond to challenges, and develop conjectures and arguments (Heinze, 2005; Hodgen & Wiliam, 2006).

In any case, teachers should know how to translate their analyses into feedback to students (KEF) – linked to the learning intentions and success criteria (KLI) – that can be used to further their learning. That is, to help them to answer the question 'How am I going?' in the feedback cycle and to envisage 'Where to next?'

The ability to interpret evidence of students' learning is also dependent on teachers' subject knowledge (Hodgen & Wiliam, 2006) and on their knowledge of how the learning of mathematics occurs (Brookhart, 2011).

5.1.1.5 Knowledge of closing the feedback loop (KCL)

If the main goal of giving feedback to students (KEF) is to help them to improve their learning, students should be given the opportunity to act upon the feedback. The only way to know whether the improvement of learning results from feedback is by giving students opportunities to make some kind of response to complete the feedback loop (Sadler, 2010). These opportunities should be given while the assignment or learning task is still relevant to the student (Brookhart, 2008; Hattie & Timperley, 2007).

In mathematics teaching and learning, this can take different forms, for example, the teacher could give students the opportunity to re-do an exercise and re-analyse their answers afterwards; or even after a test (with the possibility of changing their marks if learning improved). The latter could also be considered as a way of using summative assessments with a formative purpose which has also been proved to be fruitful in mathematics lessons (Black et al., 2003).

5.1.1.6 Knowledge of helping students use assessment information (KHS)

A major contribution of FA research is that the effects of assessment on students are powerful (Black & Wiliam, 1998). Therefore, it is one of the teacher's roles to help students to use assessment *for* learning.

When giving feedback (KEF), teachers are shaping some of their interventions to meet the learning intentions that have been made evident (KLI), but they are also implementing a very important principle of learning, recognising that students must be active in the process – learning has to be done 'by' them; it cannot be done 'for' them (Hodgen & Wiliam, 2006). Encouraging self- and peer-assessment play an essential role in this matter (Black & Harrison, 2001; Sebba et al., 2008).

In addition, especially with peer-assessment, students have the opportunity, and are encouraged, to talk to each other, which has already been proved to be an important aspect of learning mathematics (MacGregor, 2002; Hodgen & Marshall, 2005; Moschkovich, 1999, 2002). Having to express their mathematical ideas, helps students construct the mathematics language. "'Talking the talk' is an important part of learning" (Hodgen & Wiliam, 2006, p. 4) and a meaningful way of helping them to analyse and re-structure their mathematical ideas.

In regards to the six DoK, although I have given some examples of how they could be interpreted in mathematics teaching and learning, the domains themselves were mostly retrieved from the literature of FA in general. However, this compilation needed to be validated. The steps that I used to ensure validity, not only of the DoK, are explained below.

5.2 Establishing the validity of the MaTFAKi

In this study, I followed Kane's (2006) argument-based approach to validty, which means that the logic of an evaluation argument should be used to validate score interpretations. This requires the creation of two related arguments:

interpretive argument which refers to the inferences and assumptions that can be made from the responses on an instrument to an interpretation or use of its results.

validity argument which offers an evaluation of the interpretive argument to determine if it is coherent, reasonable, and plausible.

Following this approach, I have developed an interpretive argument for the MaTFAKi. The MaTFAKi was developed to measure what teachers know about FA in general and the idea of feedback in particular. The claim for the MaTFAKi was that teachers who get a high score are more knowledgeable in the domain of interest.

For the development of the validity argument, I followed a three-step process, which included Brazilian researchers and classroom mathematics teachers. In PHASE 1, researchers participated in a review of the DoK to determine if the knowledge being assessed was a good representation of the FA theory. In PHASE 2, researchers participated in a thorough review of the questions, including a validation of whether the six DoK were well-represented by the questions. Finally, in PHASE 3, teachers participated in a face-to-face review so I could analyse how they understood, interpreted, and responded to questions in the MaTFAKi.

PHASE 1 was conducted before I wrote the questions. Its results are presented in the next section. PHASES 2 and 3 were carried out after the questions were developed and therefore will be discussed in the next chapter.

5.2.1 PHASE 1: Review of the domains of knowledge

In PHASE 1, the main purpose was to validate how representative of FA the DoK chosen for the development of the MaTFAKi were, as specified in Section 5.1.1.

Eleven earlier career and senior researchers in the field of assessment[1] were asked to indicate how important each DoK were on a scale ranging from 'of no importance'

[1]Members of the research group GEPA – Grupo de estudos e pesquisa em avaliação (Group of studies and research on assessment), from the University of Brasilia – Brazil. Available at http://gepa-avaliacaoeducacional.com.br/

to ‘very important’, and also whether they felt that a substantive area (or knowledge) of FA was missing (see Appendix A.1).

To facilitate their judgement and understanding, I provided the supporting explanation of each DoK and one preliminary question as an example of how I would make use of these DoK. The review was conducted via email, and a follow-up discussion was carried out when I judged necessary.

Based on what the researchers were being asked, there were three possible outcomes: an agreement among them, a disagreement, or one (or some) of them proposing the inclusion of a new DoK. For the analysis, I followed an iterative process as shown in Figure 5.2.

If they all agreed that a DoK was ‘important’ or ‘very important’, the DoK would be accepted with no further discussion. If they agreed that a DoK was ‘of some importance’ or ‘not very important’, further discussion would be carried out. Equally, in cases of disagreement or the proposal of another DoK, I would discuss it further with them. The results can be found in Table 5.1.

Table 5.1: Results of first content validation.

Knowledge	Not important at all	Somewhat important	Important	Very important	Total
KLI	0	0	0	11	11
KAM	0	0	1	10	11
KIE	0	0	0	11	11
KEF	0	0	0	11	11
KCL	0	1	1	9	11
KHS	0	0	0	11	11

As can be seen, the only DoK that needed further discussion was KCL with one of the researchers considering it as being ‘somewhat important’. During our conversation, she explained:

> I anticipate that my answer is from the point of view of a primary-school teacher. I know your focus is the secondary school, but I understand that when feedback occurs, there is interpretation of both (teachers and students) about what is being communicated. In the case of young children, they create assumptions that will end up emerging from their written work. These are not always correct from the point of view of formal thoughts, but

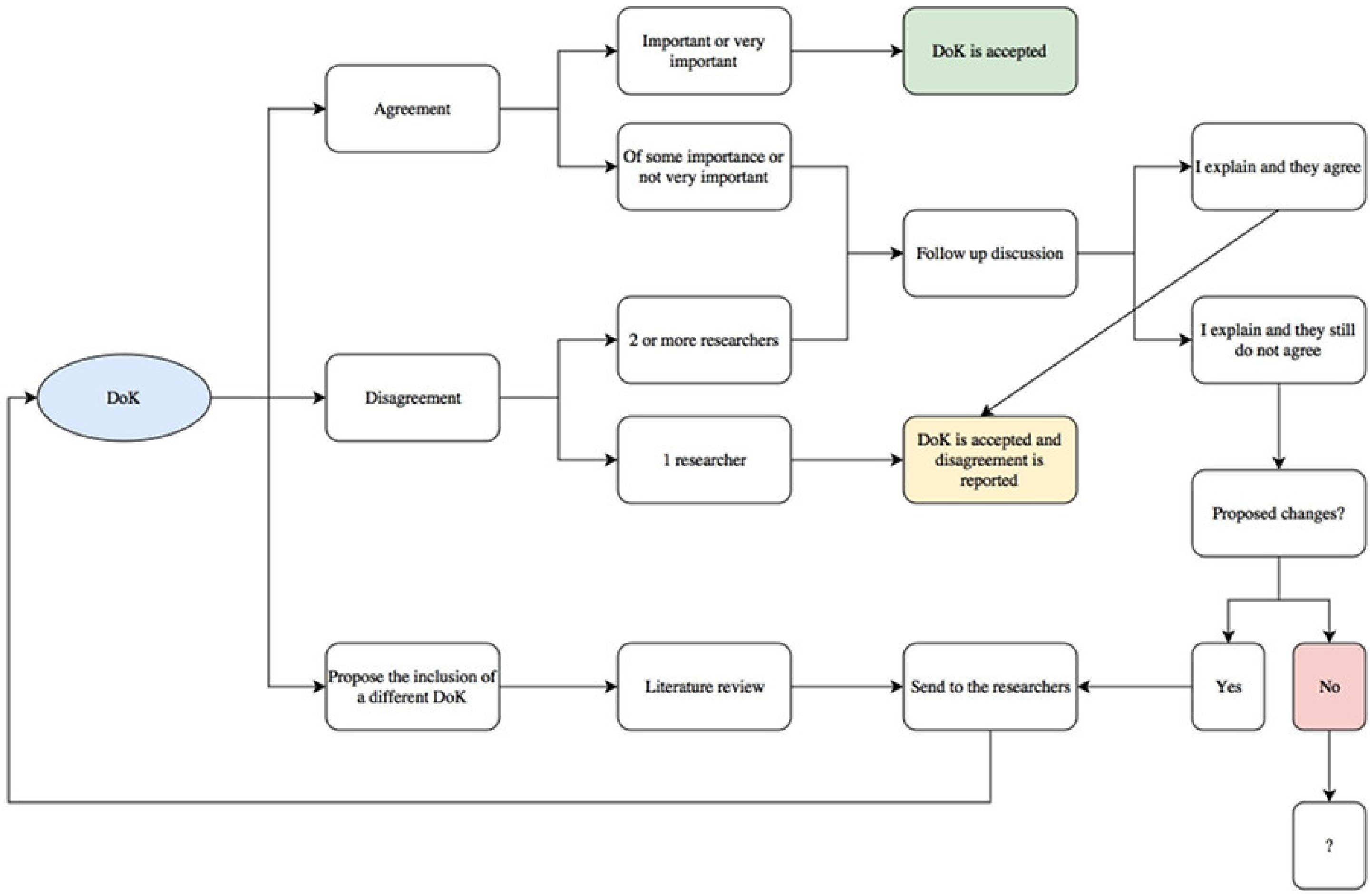


Figure 5.2: Validation of the six Domains of Knowledge.

> they are indicators of how the teacher could improve his/her own feedback and, simultaneously, of some concepts that the student is building. You introduced me to only one resolution of a student. In this context, in fact, what is being asked is if the student understands a certain sequence of calculations to solve the problem, but there are many other situations in which this could be considered.

Although her answer included very important points (e.g. different interpretations from teachers and students, or teachers using feedback to improve their practices), I realised that she judged the domain as 'somewhat important' because she thought I was only applying it to one situation (based on the preliminary question I sent as an example). Therefore, her answer showed that she was not interpreting the DoK as expected.

I then explained that this DoK was actually related to the importance of giving students the opportunity to act upon feedback in any situation (oral or written, formal or informal); that the teacher needs to give another opportunity to make sure that learning occurred based on the students' response to the feedback (Sadler, 1989). In addition, I explained that the question I had sent was just a first draft of an example of how I was intending to transform the DoK into questions, and that I was aware that "there were many other situations in which this could be considered" [quoting her]. After this further explanation, she agreed that this specific knowledge was indeed 'very important'.

Finally, one of the researchers proposed the inclusion of a different DoK: 'Knowledge of dealing with assessment information in an ethical way'. As I have already mentioned ethics as an important issue related to assessment (see Section 3.2.6), I sent his suggestions to the other researchers.

They all agreed that ethics is a very important knowledge related not only to FA, but to all kinds of assessment. However, they also argued that ethical behaviour should be present in every teaching and learning activity and not only during assessment events. Therefore, due to the feasibility of the study – a set of six DoK was already a good amount to be investigated – they agreed (including the researcher who proposed the inclusion) that this DoK did not need to be included. Furthermore, they believed that this knowledge would be embedded in the other DoK and in the answer options to the questions, and would therefore be covered in the MaTFAKi in some way.

They also made general suggestions that were very important for me to take into account when developing the questions. One suggested:

> You need to consider that, in relation to feedback, it is important that the teacher is aware that this is an external source for the student. With the teacher's help, the student will start to get used to the information provided to use it to make learning move forward, as part of the student's self-monitoring abilities. The aim of the educational work should be then to facilitate the transition from feedback to self-monitoring.

Overall, the majority agreed that these six DoK were a good start and representative of what teachers should know in relation to FA. One of them commented:

> I believe that the six domains of knowledge presented above, as necessary to the design and implementation of assessment, contemplate the elements which I judge to be essential to assessment for learning: intention, analysis of what is being assessed, tasks and assessment methods consistent with intentions, feedback, and the inclusion of students in the process.

All researchers answered within a week, and another week was necessary for the follow-ups. After these two weeks, I could conclude that the six DoK that I set out formed an acceptable framework for the development of the MaTFAKi. Taking their suggestions into consideration, I started writing the questions as I will explain in Section 5.4. Before that, I will explain how the format was defined.

5.3 Defining format

In this step of the development, the focus was on the selection of the MaTFAKi layout and the types of questions that would be included. I was very careful regarding how the MaTFAKi would be presented to teachers because having a good design is essential for producing a valid instrument (Rust & Golombok, 2009) and to the credibility of the study.

5.3.1 Type of questions

There are different types of questions that can be used in a questionnaire (dichotomous, multiple-choice, rating scales, open-ended, among others). The choice depends, among other things, on the type of construct being measured and the kind of information that the researcher wants to be able to inform with the results of the study (Cohen et al., 2011; Oppenheim, 1992).

According to many authors (e.g. Ben-Simon, Budescu, & Nevo, 1997; Furr & Bacharach, 2014; Haladyna & Rodriguez, 2013; Kline, 1986; Rust & Golombok, 2009), if the intention is to measure domain(s) of knowledge of any cognitive demand (recall, comprehension, or application), multiple-choice questions are the most suitable.

As this was the intention of this study, all questions were in a multiple-choice format with four response options each. However, instead of using the traditional multiple-choice format (with a correct answer), the questions were designed in an ordered multiple-choice (OMC) format (Briggs, Alonzo, Schwab, & Wilson, 2006).

The main reason for choosing this type of question was because it would be possible to assess partial knowledge (Ben-Simon et al., 1997). As I said before, my intention was not to criticise individual teachers by saying whether they were knowledgeable or not. My intention was to produce actionable meaningful information about teachers' knowledge of FA, and therefore this kind of question was essential to achieving that goal.

Thus, each response option was intended to represent qualitatively distinct information about teachers' knowledge. While each question contained a response option that was considered 'the best' answer, teachers were given partial credit when they selected a response that represented some understanding of what was being measured in that question. This information enabled me to collect and compare teachers' responses to understand and discern different patterns between them, consequently leading to a better view of their knowledge of FA.

In addition to the OMC questions, the MaTFAKi was also developed in a Situational Judgement Test (SJT) format. In this case, each question (or a set of questions) is embedded in a specific situation and teachers have to make an evaluative judgement based on that explicit situation. This makes it possible to relate the research-based knowledge extracted from the literature and the teacher-based knowledge that I sought to assess. Implicit in this format is that all choices have merit, but when a situation (or context) is used, one of these choices can be considered the best (Haladyna & Rodriguez, 2013). Therefore, teachers were required not only to show their knowledge, but their 'conditional knowledge', which involved judgement of how and when to do something (Bruning, Schraw, & Ronning, 1995).

According to McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001), SJTs have been used since the 1920's with the main purpose of predicting job performance. It is usually designed to assess an applicant's judgement regarding a situation encountered in the work place (Motowidlo, Dunnette, & Carter, 1990; Weekley & Ployhart, 2013). Klassen, Durksen, Rowett, and Patterson (2014), for example, used an SJT to understand non-cognitive attributes (e.g. resilience, empathy) of candidates to be selected for teacher training.

There are essentially two main purposes of SJTs: to measure knowledge or behavioural tendency (McDaniel et al., 2001; Whetzel & McDaniel, 2009). In the knowledge tendency, respondents are required to judge what 'would be the best' (or correct) option or to rate how effective each option is. In the behavioural response, participants are asked about what they 'would do' in that situation or to rate the likelihood that they would perform an action.

In this study, the former was used. First because what was being assessed was teachers' knowledge and second because the main purpose was to assess whether the respondent knew the best response to the situation. My intention was not to assess whether teachers would act like that in their lessons, but whether they 'knew how would be the best way to act' in the situation presented, which is aligned with the 'knowledge tendency' (McDaniel et al., 2001; Whetzel & McDaniel, 2009).

SJTs can have many advantages. First, it has been shown that their results show lower sub-group differences than measures of general cognitive ability, especially with regard to race difference (Chan & Schmitt, 1997; Motowidlo et al., 1990; Whetzel, McDaniel, & Nguyen, 2008). Second, as they usually describe work situations, they tend to have face and content validity (Klassen et al., 2014; Salgado, Viswesvaran, & Ones, 2001). Third, they can measure implicit traits (Klassen et al., 2014), and can therefore provide information of teachers' understanding on effective behaviour in relation to non-academic attributes (empathy, integrity).

Taking into consideration that the MaTFAKi was developed to assess knowledge of FA, particularly the feedback cycle – and that FA can take many forms – justified the use of this type of question. This is because what is 'the best' attitude in one situation, might not be considered so in another. Therefore, it would have been almost impossible to assess teachers' knowledge of FA in a decontextualised manner.

In this study, the use of SJTs also offered a fourth advantage, which was to assess situations considered very important in FA but that I would have been unable to assess if I had chosen to do observations, for example, simply because they are not common practice in mathematics lessons in Brazil (Albuquerque, 2012; Camargo & Ruthven, 2014; Camargo, 2015).

Therefore, the scenarios were based on situations that respondents were used to facing in their daily routine as mathematics teachers (and assessors), but also on essential situations of assessment being used with formative purposes which they may have not faced in their routine.

Although the MaTFAKi could be considered one type of SJT, there are three main differences that need to be highlighted, although they will be better understood when I introduce one of the scenarios in Section 5.4.1:

1. Each scenario comprised more than one question and the response options were not designed to represent the same idea across questions. Each question had its own response options that represented an idea in regards to the DoK being assessed by that question. Equally, the difference in scores was not the same across questions. That is, the difference between score 4 and 3 or between 3 and 2, for example, was not the same from question to question.

2. The 'judgement' was not used for selection, nor to predict job performance. Actually, there was no judgement being made. The MaTFAKi was developed to 'understand' teachers' knowledge of FA so the results can be used to help teachers improve their practices.

3. As I was also assessing partial knowledge, answers were ordered and teachers were given a score based on their responses.

Finally, with a more contextualised instrument, teachers' engagement in answering the questions was facilitated, which helped to reduce the amount of 'faking' or guessing (Kline, 1986; Haladyna & Rodriguez, 2013), thereby increasing the validity of the results.

Although the multiple-choice format has many advantages and fits the purpose of this study, this type of question is often criticised due to the difficulty in writing items, especially the distractors. In Section 5.4, I address these issues and what I did (or tried to do) to overcome them.

5.3.2 Layout

The MaTFAKi was developed using the web-application Qualtrics®[2]. Following Schonlau et al.'s (2002) suggestion, I used a multi-page design to reduce scrolling and to prevent teachers from going back to a question and changing their answers (see Figure 5.3 as an example).

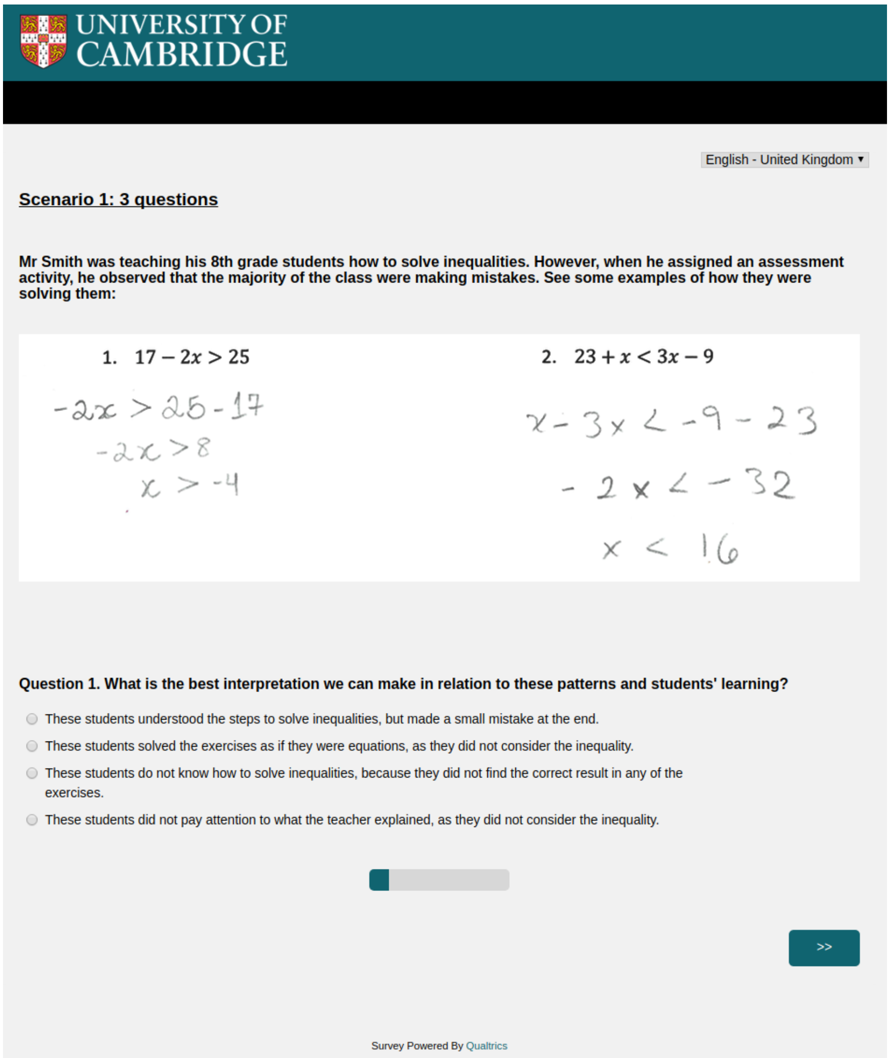


Figure 5.3: An example of a question from the online version.

This was necessary due to the scenario-based approach that I adopted, in which, sometimes, the next question would give teachers the answer (or clues) to the previous question. Therefore, they were not allowed to change their previous answer, as this

[2]http://www.qualtrics.com

would mask the results. However, in any screen, teachers could access the scenario when they deemed necessary (Figures 5.4 and 5.5).

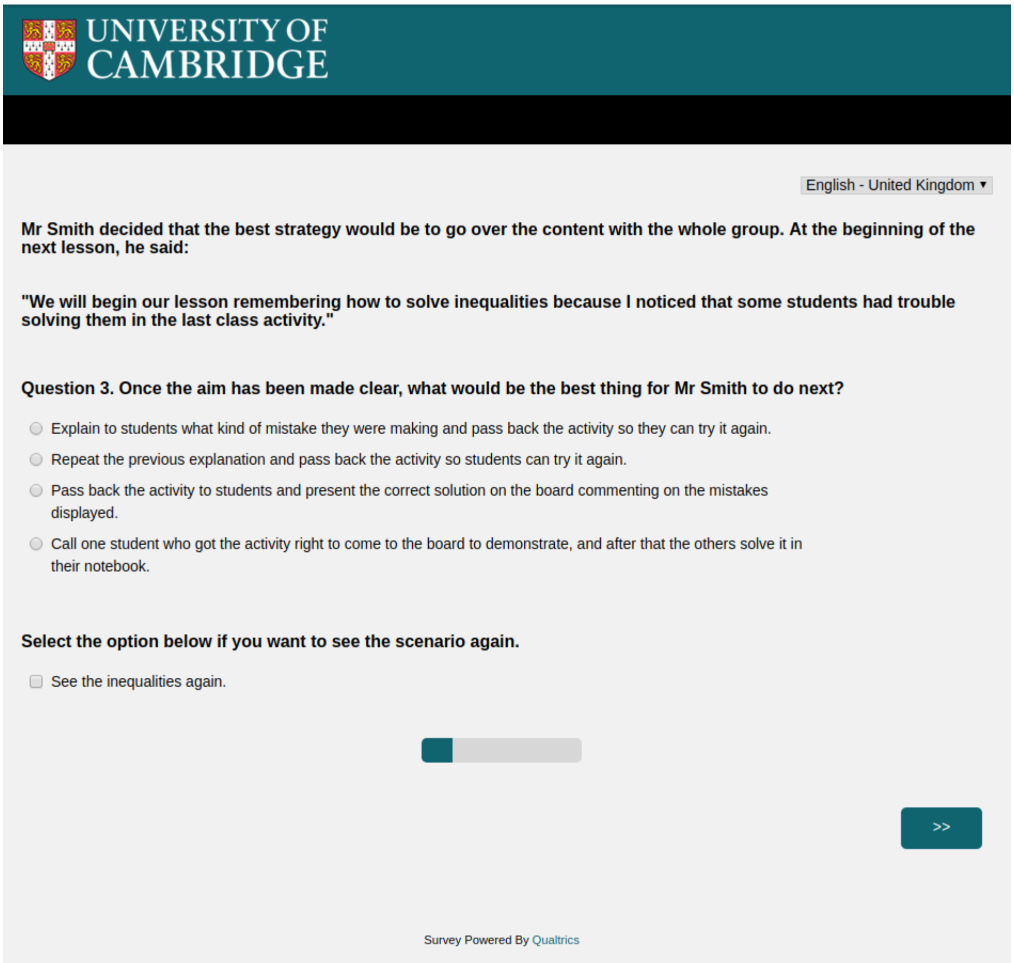


Figure 5.4: An example of how a question would first be seen - with the scenario hidden.

Additionally, all questions were compulsory – participants were unable to proceed to the next page without answering the question on the current page. Although I believe this procedure made some responders give up without completing the MaTFAKi, the type of study that I designed required this measure. As suggested by Smyth, Dillman, Christian, and Stern (2005), forced response design depends on the needs of the particular study. In this study it was important to have all questions answered so I could draw patterns among them and make reliable assumptions, which would only make sense if all teachers answered all questions.

To better understand the characteristics of respondents, the MaTFAKi also had a section for collecting demographic data: age, gender, years of experience and

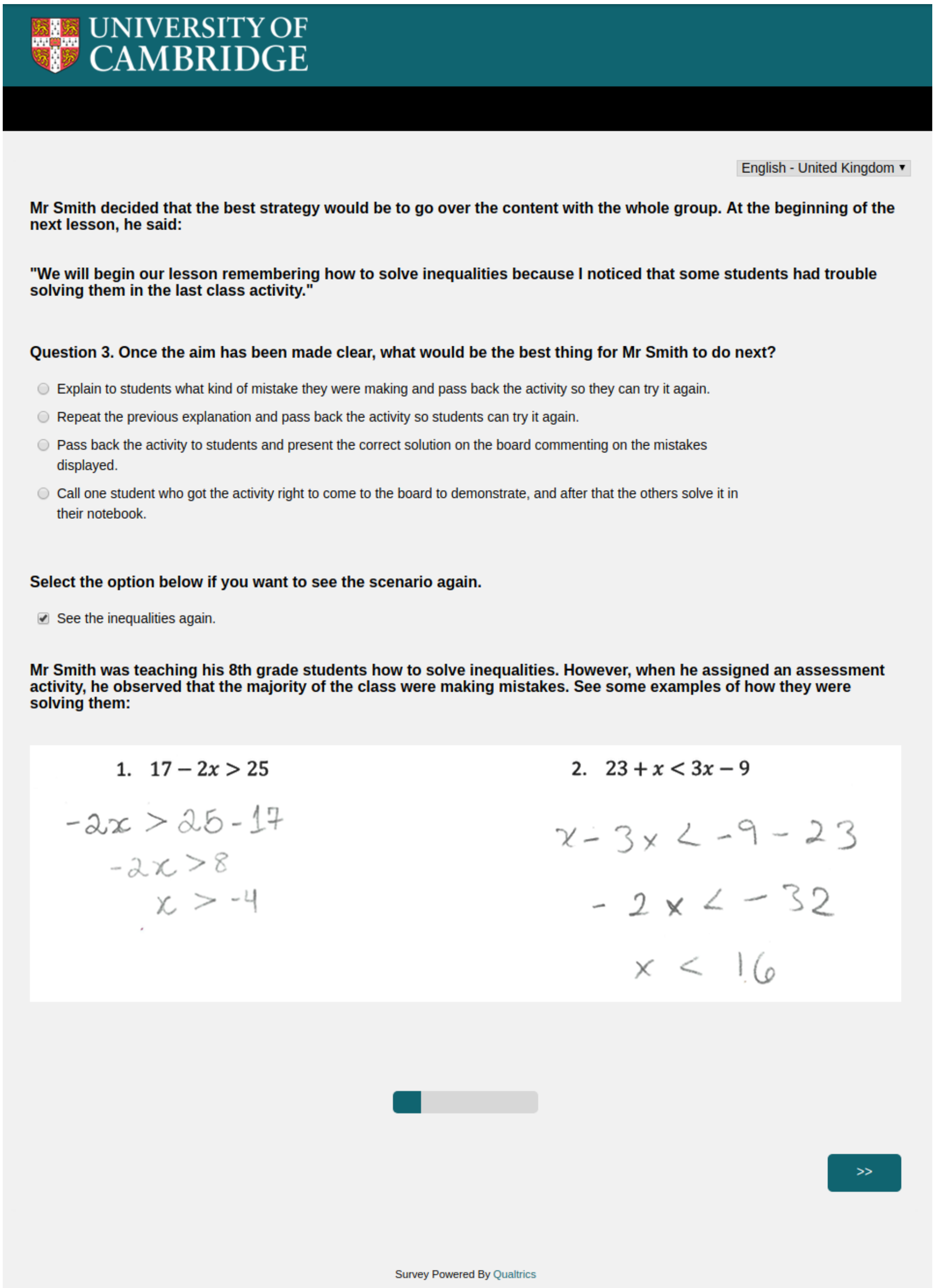


Figure 5.5: An example of a question when the respondent had chosen to see the scenario.

educational level, which was used to make comparisons and answer some of the research questions.

I also included a final question asking if teachers would like to include their contact information, for further research. As I also used this information for a prize draw (explained in Chapter 8), they had the option to choose if they wanted their contact information to be used in both or just one of them.

5.4 Writing the questions

As said before, writing items in multiple-choice format is a very hard task, which usually requires a team of experts. Having this in mind, I started by following a systematic approach for developing the questions.

First, I analysed the existing instruments and studies reported in Section 4.2 to verify whether any of the questions would be suitable or could be adapted for this study. Unfortunately, I was unable to use any of them, mostly because the majority did not include the principles of FA (they were more focused on educational assessment in general or examinations) or had elements that could not be adapted to the Brazilian context (such as the standard-based approach). No research had been conducted in Brazil.

I then started to search for research studies that could provide examples of FA being used in mathematics lessons (later on I included other subjects). Again, and I believe that it is mostly because of the (short) length of research papers which discourages presentation of such examples, I did not find a lot of help. However, those studies that were used are properly acknowledged in Chapter 7. Another source of information was websites (mostly blogs) reporting good FA practices.

Facing this drawback, I started to write classroom situations (which are referred to as Scenarios in the MaTFAKi) based on my experience as a mathematics teacher, but also focussing on the DoK that I intended to assess. The ideas and theoretical explanation that underpinned these situations are presented in Chapter 7 where I specifically detail the MaTFAKi.

Especially with regards to the answer options, I always kept in mind the guidelines for writing multiple-choice items proposed by Haladyna and Rodriguez (2013), such as: *1*) write all options following the same style, grammatical form and length; avoid specific determiners (always, never, completely, absolutely, among others), *2*) vary the location of the right answer according to the number of options, and most of all, *3*) make all options plausible: to ensure that it would not be possible to choose the best answer just by using cues from the options or stems, or to exclude any options that were clearly not the best.

In addition, the MaTFAKi was also reviewed by researchers (see Section 6.1) and classroom teachers (see Section 6.2). Initially, I developed four questions for each DoK with the intention of reducing this to three for the final version. Based on the statistical analysis that I conducted (see Section 6.3), the eventual number of questions per DoK varied. The MaTFAKi was developed in English and Portuguese simultaneously as I explain in the following paragraphs.

When planning this study, I thought that I should develop the MaTFAKi in Portuguese and then translate it into English. First because I believed that, as Portuguese is my native language, it would facilitate the creative and non-linear process of developing questions; that trying to find the correct words in English would hinder the creativity necessary for this task, making the process even more difficult. Second, because I wanted to avoid misconceptions during the translation process, mainly because the Portuguese version would be the one that would be delivered to teachers and from where the data would be generated.

However, when I was able to find some kind of example in research studies, or posts in websites, or books, they were always written in English; and when I started developing the questions, I realised that it would be very hard to follow a very structured and closed process (but I always kept in mind Haladyna and Rodriguez's (2013) guidelines). Like any other researcher, I had to adapt my initial idea and follow another process. As a result, it was easier for me to just keep the language and develop the scenario/question in English and then translate to Portuguese (Scenario 2, which was adapted from Ball et al. (2008), is an example of this). In the case when I had to develop the scenario/question from scratch (or had the help from someone from Brazil), I would start in Portuguese and then translate it to English (e.g. Scenario 3).

In any case, to guarantee that both versions were similar, in addition to my own translation, the MaTFAKi was also translated by two Brazilian Portuguese speakers, who are familiar with mathematics education and assessment research and have an extensive knowledge of English[3]. After that, the English version was also verified by an English native speaker and professional proof-reader. Both versions were also tested by several colleagues to correct typos and verify the time required for completing the MaTFAKi.

In the next section, I will present and explain Scenario 1 so the reader can better

[3]Both with a degree from a British university.

understand all the next sections.

5.4.1 Introducing Scenario 1

The main purpose of this section is to provide an example of what I mean by a scenario-based (or situational) questionnaire with OMC questions. I therefore present the rationale, scores and explanations of the questions in Scenario 1. In Chapter 7, I explain the remainder of the MaTFAKi and in Appendix B.5, I present the whole MaTFAKi as it was sent to the respondent teachers.

Scenario 1 is a short scenario which follows a 'standard' sequence. *Mr Smith* teaches (and how he taught the content is not included in the scenario), assigns some kind of activity, checks what students have learnt (Q1) and notices that the majority were having trouble. *Mr Smith* then needs to decide on the best way of communicating that to students (Q2). After that, *Mr Smith* makes his learning intentions for the next lesson clear to students, and then he has to decide how to go over the specific content again (Q3).

Therefore, in the first page, I present *Mr Smith's* intention, two examples of how the majority of students were solving the inequalities, and Q1 (see Figure 5.3 on page 100). Teachers are required to take *Mr Smith's* position and interpret students' learning (an example of KIE) based on how they were solving the inequalities in the examples presented.

The explanation of each answer option, in order of adequacy, is presented in Table 5.2.

Table 5.2: Question 1 scores and explanations

Option	Score	Explanation
B	4	Teacher interpretation of students' learning is that they did not consider the inequality. The reason is that students solved the exercises as if they were equations.
A	3	Teacher interpretation of students' learning is that they understood the steps to solve inequalities. The reason is that they had a small mistake at the end.
D	2	Teacher interpretation of students' learning is that they did not consider the inequality. The reason is that students did not pay attention to what the teacher explained.
C	1	Teacher interpretation of students' learning is that they do not know how to solve inequalities. The reason is that students did not find the correct result in any of the exercises.

Q1 is comprised of two options (Q1.B and Q1.D) where teachers identify that students *did not consider the inequality* and two where there is no interpretation in regards to making sense of how students have gone wrong in making response (Q1.A and Q1.C).

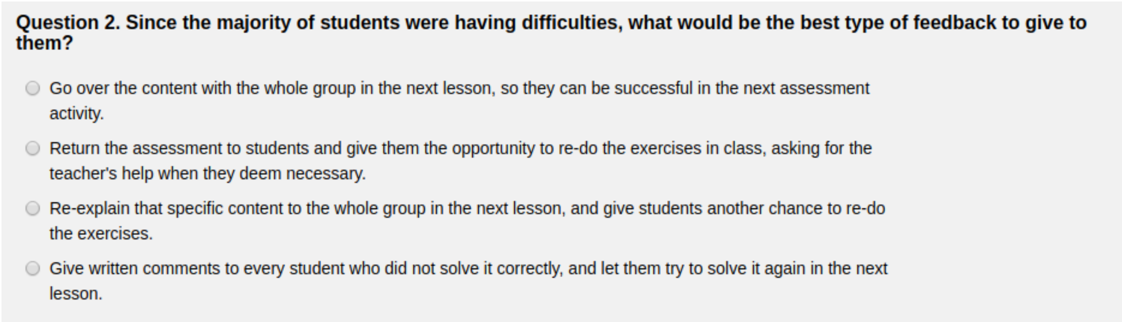
Q1.B is considered to be the best because it focusses on the process used (Hattie & Timperley, 2007) and gives a mathematical interpretation. With this interpretation, it would be possible to use the information gathered to modify instruction to facilitate improvement (Suskie, 2009) and to provide effective feedback that moves learning forward (Wiliam & Leahy, 2015).

In Q1.D, although there is also a mathematical interpretation, the reason given for students' mistakes is because they did not pay attention to *Mr Smith*'s explanation. This option was included based on results of PHASE 2, which I explain in Section 6.1. However, Bandura (1997) has also shown that teachers tend to believe that when students do well, it is as a result of teacher effort and strategies (Moore & Esselman, 1992; Ross, 1992), while when students are struggling, it is students who need to put in more effort (Gibson & Dembo, 1984). With this interpretation, it seems that the teacher might not find necessary to take action as there is no gap in learning. If students pay attention, they will be able to solve it correctly.

This is one of the reasons why Q1.A was considered to be better than Q1.D, even though Q1.A does not include a mathematical interpretation. The other is because it is possible to conclude that *they made a small mistake at the end*, whereas in in Q1.D, it is not possible to affirm that students did not pay attention to the explanation with the information provided. Q1.A is not considered as good as Q1.B because teachers do not identify what students know or do not know – they only acknowledge that there was a mistake.

Q1.C is considered to be the least appropriate option because the interpretation is that the inequality is completely incorrect because students did not find the correct final answer. Therefore, it focusses on the final result rather than the whole solution. In this case, the mistake will hardly ever be used as a learning opportunity (Bray, 2011; Heinze, 2005; Santagata, 2005).

The scenario continues with Q2 in which teachers have to judge which is the best type of feedback to be given to students (an example of KEF), taking into consideration that the *majority of students* were making the same mistake (Figure 5.6).



Question 2. Since the majority of students were having difficulties, what would be the best type of feedback to give to them?

- Go over the content with the whole group in the next lesson, so they can be successful in the next assessment activity.
- Return the assessment to students and give them the opportunity to re-do the exercises in class, asking for the teacher's help when they deem necessary.
- Re-explain that specific content to the whole group in the next lesson, and give students another chance to re-do the exercises.
- Give written comments to every student who did not solve it correctly, and let them try to solve it again in the next lesson.

Figure 5.6: Question 2.

The explanation of each answer option is presented in Table 5.3.

Table 5.3: Question 2 scores and explanations

Option	Score	Explanation
C	4	Teachers identify that when the majority are having trouble, the best strategy is to give feedback to the whole group. Teachers give students the opportunity to act upon the feedback while the feedback is still relevant.
D	3	Teachers do not identify that when the majority are having trouble, the best strategy is to give feedback to the whole group. Teachers give students the opportunity to act upon the feedback while the feedback is still relevant.
A	2	Teachers identify that when the majority are having trouble, the best strategy is to give feedback to the whole group. Teachers do not give students the opportunity to act upon the feedback while the feedback is still relevant. The focus is the next assessment.
B	1	Teachers do not identify that when the majority are having trouble, the best strategy is to give feedback to the whole group, as there is no feedback being giving at all. Teachers give students the opportunity to re-do the exercises, but leave to students' interest to look for help.

Knowing how to give effective feedback includes knowing the audience to which the feedback is intended (Brookhart, 2008). In Q2.A and Q2.C, teachers recognise that, in this case, oral feedback to the whole group would be the best strategy. However, the purpose of giving the feedback is different in the two options. In Q2.C, the focus is on students' learning at that moment, which is aligned with the ideas of FA (Black & Wiliam, 1998; Harlen, 2012; Bennett, 2011). The teacher is using the information to determine what would be the most appropriate next steps for those students (Hodgen & Wiliam, 2006). In addition, the teacher is also giving students the opportunity to re-do the exercises after the feedback has been given, which is an important part of the feedback cycle (Sadler, 2010).

In Q2.A, on the other hand, the focus is on the next assessment. Therefore, it has a formative approach (the teacher is taking action based on assessment results) (Black & Wiliam, 1998; Harlen, 2012) with a summative intention (to be successful in the next assessment), rather than referring back to the assignment or learning task whilst it is still relevant to students (Brookhart, 2008; Hattie & Timperley, 2007).

Even though, in Q2.D, teachers do not identify that feedback to the whole group would be the best strategy, feedback is being given and, as in Q2.A, the teacher also gives students the opportunity to act upon the feedback. In this case, the teacher would spend a lot of time writing feedback to several students, wherein the same goal could be achieved by approaching the content with the whole group (London & Sessa, 2006).

In Q2.B, although the teacher gives students the opportunity to re-do the exercises, no feedback is being given. In this case, students would not be able to know 'Where they are?' and there is no indication of what they should do – there is no recipe for improvement (Wiliam, 2011) to answer the question 'What next?' in the feedback cycle (Hattie & Timperley, 2007). This was discussed in depth with the researchers and all agreed that it is better to give feedback focussing on the next assessment than no feedback at all. That is why Q2.A was judged to be better than Q2.B.

The scenario continues on the next page with a short sentence explaining that *Mr Smith* decided *to go over the content with the whole group*, and showing how he made clear his intentions to students in that lesson. Next, Q3 is introduced (see Figure 5.4 on page 101).

Teachers were required to choose the best way of going over the content with the whole group. In this case, what was being assessed was whether teachers recognised the importance of letting students act upon the feedback being given (Sadler, 2010; Shute, 2008).

Therefore, in Q2, the teacher had to judge 'which type' of feedback would be better, whereas in Q3, they had to judge 'how' this feedback should be given to offer students the opportunity to identify what was missing and subsequently act upon the feedback (assessing the KCL).

Table 5.4 shows the scores and explanations of each option in Q3.

Table 5.4: Question 3 scores and explanations

Option	Score	Explanation
B	4	The teacher goes over the content of inequalities without specifically focussing on the mistakes. Students have the opportunity to try again after the explanation.
A	3	The teacher focusses their explanation on the mistakes displayed. Students have the opportunity to try again after the explanation.
C	2	The teacher focusses on showing the correct solution, while explaining the mistakes displayed. It is not clear if students will have the opportunity to try again or just copy the correct solution from the board.
D	1	The teacher does not provide any explanation. It is left to a student that got the exercises right. The other students will solve the exercises in their notebook after that.

Q3 is comprised of two options (Q3.C and Q3.D) in which the activity is passed back to students first, and two (Q3.A and Q3.B) in which some kind of feedback is given before passing the activity back.

Q3.B is considered to be better than Q3.A because the teacher would go over the content with students (focussing on the content of inequalities) who, instead of focussing only on the mistakes displayed, would have the opportunity to try by themselves and check whether or not they understood the content after being given feedback (Sadler, 2010). In a way, Q3.B encompasses Q3.A. While going over the content, the teacher can emphasise some mistakes that students made; not necessarily telling them that those were their mistakes, to make them realise for themselves. This was also discussed in depth with the researchers and in the Brazilian context, unfortunately, if the teacher shows the mistake, when they pass back the activity, most students will correct what was wrong in their first attempt (usually just copying from the board) and move on.

In Q3.C and Q3.D, equally, there is nothing else to be done afterwards, because the job has already been done for students (Brookhart, 2008). Q3.A is considered to be better than Q3.C because the teacher passes back the activity after the explanation. Q3.C is considered to be better than Q3.D because the solution comes from the teacher and some explanation is being given (*commenting on the mistakes displayed*), whereas in Q3.D, students will only visualise a correct solution from a colleague. Again, it is likely that students will just copy the correct solution from the board, with no reflection on how to improve (Castelló & Monereo, 2005).

Considering the scenario as a whole, it assessed whether respondents were able to notice that oral feedback is being given to the whole group based on what the teacher noticed through an activity prepared for that purpose. The learning intentions are made clear to students and the teacher gives them the opportunity to act upon the feedback.

5.5 Summary

In this chapter, I have explained and reflected on the steps that I followed while writing and before evaluating the questions that would compose the final version of the MaTFAKi. I also presented one of the scenarios so the reader could better visualise the MaTFAKi and understand how the scenarios and OMC questions were structured.

Now that I have exemplified how one scenario 'looked', in the next chapter I describe and explain all the phases that I followed to evaluate the questions and to ensure the validity and reliability of the MaTFAKi in assessing Brazilian mathematics teachers' knowledge of FA.

CHAPTER 6

EVALUATING THE MATFAKI

The previous chapter detailed PHASE 1: the review of the DoK, which was conducted before writing the questions. This chapter explains the evidence-based processes followed to evaluate the questions.

Section 6.1 explores PHASE 2: the review with two teams of Brazilian researchers. Section 6.2 details PHASE 3: the face-to-face review with FD classroom mathematics teachers. Section 6.3 presents the results of PHASE 4: a field test with a larger sample of mathematics teachers.

6.1 PHASE 2: Review with researchers

PHASE 2 was divided into two sub-phases. SUB-PHASE 2A comprised a thorough review of the questions. SUB-PHASE 2B checked the alignment between each question and the DoK it was developed to assess.

6.1.1 SUB-PHASE 2A: Review of the questions

SUB-PHASE 2A was conducted twice. It started with a version containing 13 questions embedded in six scenarios (Appendix B.1). Later, it was done with a 24-question version, also in six scenarios (Appendix B.3). On both occasions, I had three goals: *1*) to have a well-grounded and justified order of answer options for each question; *2*) to verify whether the options included were representative of what teachers might answer; and *3*) to analyse the alignment between each question and the DoK it was supposed to measure.

To achieve the first two goals, I asked a team of researchers from the University of Brasilia to answer the MaTFAKi. However, instead of asking them to indicate the best answer, they were required to individually rank the options to indicate the order of appropriateness of the answers, based on the situation in which the question was embedded, underpinned by their expertise. In addition, they were asked to comment or make notes for themselves on the items, for example, if they found any problems when answering them, or thought that there was an option that should be rephrased or replaced for another that teachers would be more likely to answer.

I asked them to rank the options because, when using OMC items, it is common practice to include misconceptions relating to the specific knowledge being assessed (e.g. Briggs et al., 2006). I developed the options based on my own experience as a FD teacher, and on studies available on FA and mathematics teaching and learning. As these studies did not describe teachers' misconceptions about FA, the researchers' responses (supported by their professional working knowledge of the relevant studies) could provide a basis for validating my options.

I decided to ask the researchers to analyse the MaTFAKi before discussing it with me, so they could critically analyse the questions and elicit problems. If I had just introduced them with my rationale, they might have tended to agree with me and I would not have got the kind of results that I was expecting in SUB-PHASE 2A. Therefore, asking researchers in the field, who also have a great knowledge of the system, helped to ensure that the options were firmly grounded in authentic patterns of responses rather than steered by 'my' a priori model. The answers given by each researcher, for both the 13 and the 24-question versions, before we discussed and improved the questions and options, can be found in Appendix D.1.

After that, I met them for a follow-up discussion. In this meeting, we discussed the results from their ranking so I could better understand their rationale for choosing the order of the options, and to clarify any misunderstanding they still had in relation to the items and the purpose of the MaTFAKi. After making all the necessary changes, we also agreed on the best order for the four options in each question. The meeting was video and audio recorded and transcribed verbatim. Their suggestions and arguments are embedded in the MaTFAKi rationale presented in Chapter 7.

Although I had three goals with SUB-PHASE 2A, the discussion with the researchers turned out to be much more fruitful than I initially envisaged. In the next section, I explain and exemplify some of the issues that were raised, and the subsequent changes I made. I will start by including an example of a full narrative about the process of reviewing one question to illustrate our conversation. After, I summarise the same type of issues that arose in relation to different questions.

It is important to clarify, however, that it is not my intention to show every single detail of everything that was changed based on the researchers' suggestions (for the purposes of comparison[1], the different versions can be found in Appendices B.1 to B.4), but to illustrate the importance of SUB-PHASE 2A in producing the MaTFAKi in a way that would encompass the varied ideas of FA embedded in the Brazilian, and particularly the FD context.

6.1.1.1 A full narrative of how our conversation happened

In this section, the aim is to present a full narrative of the process used to review the items in the MaTFAKi. I will show the suggested changes and some important comments that the researchers made in relation to the first version of Q1 (Appendix B.1). For every question, we would start by improving the options and the scenario if necessary, and then we would discuss and agree on the best order for the responses, considering the revised options.

Table 6.1 shows how many researchers chose each Q1 option *before* our meeting. The numbers from 4 to 1 represent the scores for each option, from the most to the least appropriate.

Table 6.1: Options ordering - Q1

Question 1				
Option	4	3	2	1
A	2	5	4	0
B	9	1	1	0
C	0	5	6	0
D	0	0	0	11
Total	11	11	11	11

As can be seen, there were nine researchers who judged that Q1.B should be accorded score 4 and all agreed that Q1.D should be accorded score 1. Their opinions in regards to Q1.A and Q1.C were divided.

[1]The questions numbers and the order of the options changed across versions. The correspondence between them can be found on Table 11.2 on page 289.

We started our discussion with Q1.D, which they considered to be the most problematic. According to the researchers, there were three main issues.

First, they were certain that the word *anything* was the reason for making everyone judge this option to be the least favoured. They suggested a more neutral option that would somehow 'blame' students' behaviour, as this is a common justification used by Brazilian teachers.

Second, the second part (*because they should have kept the x in the right-hand side of the > instead of bringing it to the left side*) did not explain the first part or the students' mistake in this case.

Third, this option was much longer than the other three, which was not compliant with Haladyna and Rodriguez's (2013) guidelines: "Keep the length of options about equal" (p. 91). Including their suggestion of 'blaming students' would also solve this problem. Table 6.2 shows the two versions.

Table 6.2: Different versions of Q1.D.

Previous version	Revised version
These students did not understand anything that the teacher taught about inequalities because they should have kept the x in the right-hand side of the > instead of bringing it to the left side.	These students did not pay attention to what the teacher explained, as they did not consider the inequality.

Regarding Q1.B, one researcher pointed out:

> What worries me is the part *but they did not take the > sign into account*. I believe that the essential difference in conceptual and representational terms between equations and inequalities lies in the solution. In the equation it is a number, a point, whereas in inequalities it is an interval. So, I don't think it is just the 'greater than' sign, maybe you can keep it in parenthesis, but I think you should change it to inequality, because the sign is not the essence here. The problem is actually *not taking the inequality into consideration* [my emphasis].

In addition, they *1)* indicated a problem with the verb tense and suggested changing *are solving* to *solved* because "they are not solving anymore, it is already solved" *2)* suggested to remove the *because they did all the steps correctly* part so that I would not have a *because* and a *but* in the same sentence, and it would make the option more closely aligned to the scenario since the students did not solve all the steps correctly – otherwise they would have found the correct solution. Table 6.3 shows the different versions.

Table 6.3: Different versions of Q1.B.

Previous version	Revised version
These students are solving as equations, because they did all the steps correctly, but did not take the > sign into account.	These students solved as if they were equations, as they did not consider the inequality.

They also suggested changes to Q1.A as mentioning that students had made a *very simple mistake at the end* could render Q1.A implausible. First because of the word *very*, and second because the mistake was not *at the end*, but during the solution. Going further in the discussion, we agreed that the best way of solving this problem would be to take out the word *very* in the option, and to change the student solution in the scenario to show that the mistake was indeed *at the end*. Therefore, the change in Q1.B was mainly changing *a very simple mistake* to *a small mistake* (Table 6.4) provided that the inequalities solution would also be changed (Scenario 1, Appendix B.4). They also suggested to change *understand all the steps* to *understand the steps*, because if students had understood *all* the steps they would not be making mistakes.

Table 6.4: Different versions of Q1.A.

Previous version	Revised version
These students understand all the steps to solve inequalities, but they had a very simple mistake at the end.	These students understood the steps to solve inequalities, but had a small mistake at the end.

Although I did not change anything in Q1.C, at this stage, it is important to mention that some researchers flagged an issue with including *students do not know* how to solve inequalities. For them, it makes this option very strong in terms of language, because it specifically says that students *do not know* how to solve the inequality and in their opinion this was not the case. However, others argued that this part made the option very interesting, and as one researcher pointed out, it would show that teachers choosing this option believe that "teaching and learning mathematics is about getting the correct results, and therefore, if the student did not find the correct results, they do not know how to solve inequalities". When I explained that that was my idea, everyone agreed that I should keep it the way it was. They also reiterated that this was the reason why we should consider Q1.A as a better option than Q1.C. In addition, if we removed the first part and kept only where it said that the students did not find the correct results in any of the exercises, it would make this option completely correct, which was not the purpose.

Based on all these changes and arguments, we agreed that the order, from the most to the least adequate would be: B, A, C, D[2].

In addition, we also agreed that with these changes Q1 would have two options, A and B, which imply that students *do* understand or know the content, and two options, C and D, in which students *do not* understand or know the content.

Although Q1.A implies that students understand the steps, it only acknowledges that there was a mistake with no interpretation involved, whereas Q1.B focusses on the process used and the interpretation is that their mistake was not taking the inequality into consideration.

In Q1.C, students are seen to not understand the content because they did not find the correct solution, whereas in Q1.D, their non-understanding is because they did not pay attention to the teacher's explanation.

This review was conducted for every question in the 13-question version of the MaTFAKi in the same day. The meeting was held at the University of Brasilia and lasted four hours.

In the next sections, I present other issues raised during this meeting, and the changes made based on our conversation. Many of the changes presented here could be included in various categories.

6.1.1.2 Finding solutions for issues with the language

When reviewing the questions with the Brazilian researchers, they identified many issues concerning how the options were phrased: some related to a different interpretation of the same word or concept, some to problems with the translation, or some to verb tenses, among other things. Therefore, this section refers to changes made to improve the clarity of the questions, without affecting content. This also included correcting spelling, grammatical, capitalisation, and punctuation errors.

Q2.B and Q2.C (Appendix B.1) are good examples of a *different interpretation of a concept*.

[2]Refer to the version in Appendix B.2, as I changed *the order of presentation* of the options afterwards (Appendix B.4) to comply with Haladyna and Rodriguez's guidelines of varying the position of the correct answer.

The researchers suggested replacing the word *practise* (praticar in Portuguese) to *re-do* (refazer) and the word *re-teach* (re-ensinar) to the phrase *go over* (retomar). The reason for the latter was because they believed that you cannot re-teach something; that teaching and learning is a process and therefore they come together. It is not possible to say that you will *re-teach* something because it suggests that teaching was already completed, thereby implying that the problem lies with the students, who did not actually learn. However, as it is a process, you can only say that something was taught if learning also occurred.

In fact, this was the first thing that the researchers wanted to comment on in relation to this question. Some explained that the word *re-teach* was the reason why they did not judge Q2.B as the best option (Appendix D.1). According to them, it was a conceptual problem, whereas in Q2.A there was no such problem.

There were also some issues with translation, excluding different interpretations as discussed above. For example, when translating *Mrs Brown*'s feedback (Q6, Appendix B.1), I translated *correctly solved* as *corretamente resolveu*. However, in Portuguese, the position of the words needs to be inverted (*resolveu corretamente*) and the researchers were very helpful in identifying these problems. Although it could be considered a simple mistake, they may prove harmful to respondents, causing them to score lower than they would have, had the mistake not been there. In addition, they can diminish credibility, consequently affecting the validity of the scores (Haladyna & Rodriguez, 2013).

6.1.1.3 Giving ideas for additional questions

One concern raised during our discussion was in regard to the number of questions. Even though I have mentioned that the version of the MaTFAKi that we were analysing at that time was the first version and that it was already in my plans to develop additional questions, the researchers felt it necessary to mention that during our discussion.

However, instead of just pointing out this concern, they also gave me some suggestions of how I could develop new questions or use those I already had to develop additional questions that would work as a follow up to those.

That was the case, for example, with *Miss Forbes*' comment in Q8 (Appendix B.1). They believed that the idea of helping students use feedback information was so important that I should turn it into a separate question.

Furthermore, turning Q8.C into a separate question, and creating a new Q8.C, solved another problem with Q8. During our discussion, everyone agreed that *Miss Forbes*' opinion did not exclude *Mrs Johnson*'s opinion, as both were plausible explanations for why students kept making the same mistake after receiving the teacher's written feedback. The changes are shown in Table 6.5.

Table 6.5: Different versions of Q8.C[3].

Previous version	Revised version
In my opinion, this happened because the students are not used to using the teacher's written feedback. You need to help them to use feedback for their own learning.	The students were unable to identify their mistake through your comments. You should have been more specific that the problem was considering the negative values.

In this revised version, the option would blame the feedback itself. The original Q8.C was then used to develop Q12 (Appendix B.4).

Q21 (Appendix B.4) is another example of a question generated from the researchers' ideas. During our discussion, they mentioned that no question dealt with the idea of being able to articulate *clear success criteria*. Although they did not help with the writing of the question itself, they suggested that a good idea would be to frame it in a way that would highlight the importance of devising the criteria together with students (Orsmond et al., 2002; Wiliam, 2011).

6.1.1.4 Including more of the Brazilian context into the MaTFAKi

Some changes the researchers suggested included adding more about the Brazilian context into the MaTFAKi.

One of the proposed changes was to replace all instances of *homework assignment* to an *assessment activity*, an *assessment assignment*, or even just *an assessment* because, although using homework assignments as assessment is a common practice of mathematics teachers in Brazil (Albuquerque, 2012; Camargo & Ruthven, 2014), teachers usually do not go over the assignments individually. This is especially the case in secondary schools, where mathematics teachers have approximately 240 students in total. They believed that making this change would also make the scenario more realistic to teachers.

[3] Q11 in Appendix B.4.

Another important change was in relation to the transition from Q12 to Q13 (Appendix B.1). They agreed that Brazilian parents would come to the school to try to understand why a teacher would change their child's marks, such as in Q12, but they considered that teachers would hardly ever send an email to parents to explain it, and that it was even less likely that parents would answer asking how they could help their children. Therefore, when I suggested to change it to the teacher giving advice to students instead of their parents, they agreed that that was actually a quite common practice in our mathematics lessons and would smartly solve the problem (Q18, Appendix B.4).

6.1.1.5 Changing how the scenario was presented to teachers

The researchers were not focussing only on the questions or answer options. They also suggested changes to how the scenario was presented to teachers. In some cases, the changes were about making the scenario more aligned with the answer options, which was the case with the inequalities in Scenario 1 (see Section 6.1.1.1). In others, the changes were about making it clearer to teachers, as with Scenario 4.

They suggested rephrasing the exercise on the board as it was not clear that the discount should be calculated each month *on top of* what was left from the previous month. To make it clearer, they suggested including the word *successively*[4].

Yet, in some cases, there were so many changes necessary, that the best solution was to delete the whole scenario. This was the case with the first version of Scenario 2 (Appendix B.1). During the meeting we agreed that the scenario included important elements, but the question embedded in it was not actually assessing any DoK and therefore needed adjustments that would change it substantially.

We then agreed, instead of adapting the current scenario, I should develop one encompassing the idea of peer-assessment, as this has been proven to be an essential element of FA (Brookhart, 2011; Sadler, 2010; Wiliam, 2011) and uncommon practice among mathematics teachers in Brazil (Camargo & Ruthven, 2014). I accepted their suggestion and developed Scenario 4 (Appendix B.4).

[4]The different versions can be found in Appendices B.1 and B.4 respectively.

6.1.1.6 Changing the questions and options to make sure they expressed a specific idea

Although all the options were written to *tell me something about the teacher who chose that option* and therefore had a specific underlying idea, this section presents some changes the researchers suggested to ensure that options or questions were expressing the intended idea.

An example is Q3.B (Appendix B.1). Even though it would contradict Haladyna and Rodriguez's (2013) guidelines of writing all options similar in style, the researchers suggested that it would be interesting to keep the *I will show you* at the beginning of the option to give an idea of ownership – of the teacher being the owner of knowledge and the one responsible for 'transmitting' it to students. In the other options, the teacher includes students in the process: *we will go over*, *We are going to start* and *we will correct*.

This type of change also happened in Q13.B (Appendix B.1). We purposefully chose to change *students should ask for their parents' help* to *students should look for help* so respondents would be able to interpret *the help* anyway they wanted: teachers', parents', a colleague's, etc. In addition, this would make the question more open, and at the same time more plausible, as it would remove the specificity to parental help.

Another example was Q2.C (Appendix B.1). To give an idea of 'teaching to the test', they suggested *[...] so they can be successful in the next assessment.* This could also be considered as *including more of the Brazilian context into the MaTFAKi* since in Brazil the 'future' of each student is dependent only on teacher-made assessment and therefore it is common practice to focus on getting better results on tests.

6.1.1.7 Focussing on the guidelines of writing good questions

This section includes the researchers' suggestions that were about changing the questions to comply with Haladyna and Rodriguez's (2013) guidelines of writing good multiple-choice questions. I present some examples below, keeping the same numbers as presented in Haladyna and Rodriguez (2013, p. 91).

Guideline 10. Minimize the amount of reading in each item: This happened with many questions. A good example was Q3 (Appendix B.1). The researchers suggested removing the final part of all options *I want to give everyone the opportunity to learn what was taught in the last lesson*, *and I want you all doing it* *correctly*, etc. for two reasons. First, to minimise the amount of reading. Second, they felt they were too romantic and at the same time, as "every teacher wants all their students learning...", they were not adding anything important to the purpose of the question.

For the type of questionnaire I was producing, with sometimes necessary long scenarios, this was a very important guideline to focus on. As the MaTFAKi had to be wordy in many aspects, I had to make sure that I could reduce the word count as much as possible to diminish the amount of time teachers would spend answering it and, as a consequence, increase the number of respondents. In addition to their suggestions, I also re-read the MaTFAKi many times focussing solely on this guideline and cutting all unnecessary words.

Guideline 11. State the central idea clearly (and concisely)[5] in the stem and not in the options: That was the case with Q2.A (Appendix B.1). The researchers suggested taking the word feedback out of Q2.A and putting it in the stem, as they believed that leaving *feedback* in the option would make teachers choose it solely because giving feedback is seen as culturally and pedagogically desirable. As only Q2.A contained the word *feedback*, it would pull the majority of answers and therefore be considered a tricky item (**Guideline 6**) since it was not considered to be the best answer in this case.

Therefore, the stem was changed to *Since the majority of students were having difficulties, what would be the best type of feedback to give to them?* and the answer option to *Give written comments* (Q2.D, Appendix B.4). This change would diminish the chances of teachers choosing this option based on social desirability. In addition, as the initial purpose of the question was to measure KEF, including the word feedback in the stem would be more aligned with the idea of stating the central idea clearly in the stem and not in the options. The word *comments* was also purposefully chosen to follow **Guideline 20c. Avoid clang associations, options identical to or resembling words in the stem** (Haladyna & Rodriguez, 2013, p. 91).

Guideline 20a. Keep the length of options about equal: That was the case with Q2.C (Appendix B.1). In addition to just being a shorter version of Q2.B, which

[5] I included *concisely* in parenthesis because, although this is the guideline, there were some situations in which it was not possible to be concise as I explained in Guideline 10 above.

would give clues to the teachers, it was much shorter than the other options.

In some cases, however, it was not possible to follow this guideline. Therefore, in these situations, I accepted one of the researchers' advice (who has experience in developing questions for external examinations in Brazil) and followed the Brazilian recommendation – *asymmetry*: when it is not possible to maintain the same length throughout, they should be ordered from the longest to the shortest or shortest to the longest. That was the case with Q9[6] (Appendix B.1).

Guideline 20f. Keep options homogeneous in content and grammatical structure: In addition to the changes already mentioned, the researchers suggested something else for Q3 (Appendix B.1). They recommended including *I noticed* at some point in all options to keep the same structure (Q3, Appendix B.4). All options would start with the teacher saying what she would do in that lesson, then follow with *because I noticed*, followed by *what* the teacher had noticed (the students had trouble, the student got it wrong, etc.).

As a result, all answer options would also be giving the idea that the teacher was using assessment *for* learning, emphasising the idea of FA in the MaTFAKi. Based on the assessment activity, the teacher *noticed* that some students were having trouble solving inequalities. Therefore, she would go over that specific content, so they could improve their learning and achieve the expected goal.

In every change we made, we would always bear in mind the importance of keeping all options plausible (Guideline 21).

As explained at the beginning of the chapter, SUB-PHASE 2A was done twice. In the second time, because the MaTFAKi was longer, I met with the researchers to discuss each scenario separately. In addition to checking the content of each question, I also validated their alignment with the DoK, which I explain next.

6.1.2 SUB-PHASE 2B: Alignment between questions and DoK

For SUB-PHASE 2B, I re-sent the MaTFAKi to the researchers and asked them to complete the form in Appendix A.2. The purpose was to check the degree to which they felt that the question-to-DoK alignment was appropriate.

[6]This question became Q22 in the field test (Appendix B.4). Q21, in the same appendix, is another example of *asymmetry*.

Although I decided not to ask them to explain their rationale in the form, I made clear in the consent letter that if I had questions about their answers, I would contact them again for a follow-up discussion.

Again, all 11 researchers from PHASE 2 participated in this stage. As agreed in the email, one week later, all researchers had already returned the forms with their judgements. Their answers are presented in Table 6.6.

Table 6.6: Results of the item-DoK alignment review.

Questions	Item-DoK alignment				
	Very low	Low	Moderate	High	Very high
1	0	0	0	0	11
2	0	0	0	1	10
3	0	0	0	0	11
4	0	0	0	2	9
5	0	0	0	0	11
6	0	0	0	0	11
7	0	0	0	0	11
8	0	0	0	2	9
9	0	0	0	0	11
10	0	0	0	0	11
11	0	6	5	0	0
12	0	0	0	3	8
13	0	0	0	0	11
14	0	0	0	10	1
15	0	0	0	4	7
16	0	0	0	2	9
17	0	0	0	0	11
18	0	0	0	2	9
19	0	0	0	0	11
20	0	0	0	1	10
21	0	0	0	0	11
22	0	0	0	0	11
23	0	0	0	0	11
24	0	0	0	4	7

As can be seen, except from Q11, all items were rated as being highly or very highly aligned to the DoK they were developed to measure. Therefore, I took no action at this stage. Later, however, these results were used in addition to the statistical analysis when deciding which items to keep in the final version (see Section 6.3.2).

The sections above show different aspects in which the researchers helped me to produce a better MaTFAKi; one more aligned with the Brazilian context and what teachers face in their daily routine as assessors. However, even though all researchers who participated in PHASE 2 had experience in teaching mathematics, and were in constant contact with teachers, the only way of checking whether the MaTFAKi was indeed assessing what it was developed to assess, would be by analysing how classroom teachers understood, interpreted, and responded to questions in the MaTFAKi. Therefore, after incorporating the changes suggested by the researchers, I conducted a face-to-face review with FD mathematics teachers.

6.2 PHASE 3: Review with teachers

As explained in Section 5.2, the primary goal of PHASE 3 was to understand how teachers read and made sense of the MaTFAKi, to ensure they interpreted the information as expected.

In the first time that PHASE 3 was conducted (PHASE 3A), the six mathematics teachers below[7] were asked to answer the MaTFAKi while I observed them. All conversations were audio recorded and transcribed verbatim. The version that they answered at this stage can be found in Appendix B.2.

Marjorie has been teaching mathematics in FD state schools for 24 years. At the time she answered the MaTFAKi, she was acting vice-principal of her school, and therefore did not have her own classes[8]. She has a teaching qualification in mathematics and a post-graduate certificate in 'Science, mathematics and the use of ICT'. Her school is located in an administrative region of the FD[9].

Jordan has a total of nine years' experience in teaching mathematics, four in FD state schools. He has a teaching qualification in mathematics and was finishing another in physics. He was teaching grades 6 and 7 (11 and 12 years-old students) when

[7] All names are fictitious to protect the participants' identity.

[8] In the FD, when a teacher is unable to work – due to family or health issues – usually it is the principal or vice-principal who acts as the cover/supply teacher, and therefore Marjorie has to teach quite often.

[9] The FD is comprised of 31 administrative regions (see Figure 2.2).

he participated in this review. His school is located in the same administrative region as Marjorie's.

Cathy has been working in independent and state schools in the FD for 26 years. She has a degree in science and mathematics with a teaching qualification in mathematics, a post-graduate certificate in 'Mathematics for Teachers' and a Masters degree in mathematics. She is currently working as a teacher trainer for mathematics teachers from FD state schools. Her school is located in Brasilia.

Amelia has been working as a mathematics teacher for 10 years, with five in FD state schools as a temporary teacher[10]. At the time she participated in this review, she was teaching 10 classes of the equivalent to AS-level students. She has a teaching qualification in mathematics and a post-graduate certificate in 'Mathematics for Teachers'. Her school is located in one of the administrative regions.

Martin has 25 years' experience in teaching mathematics in independent and state schools in the FD. He has a teaching qualification in mathematics and a post-graduate certificate in 'Teaching Higher Education'. He was teaching grade 9 (14 years-old students) when he participated in PHASE 3. His school is located in Brasilia.

Max has been teaching mathematics in FD state schools for 22 years. He has a degree in mathematics and science, with no teaching qualification, and a post-graduate certificate in 'Environment issues and energy saving'. At the time he participated in this review, he was teaching grades 8 and 9 (13 and 14 years-old students). His school is located in one of the administrative regions.

These teachers were asked to indicate any problems that they found while answering the MaTFAKi and explain why they chose a specific option instead of another. Therefore, my specific goals with PHASE 3 were to understand whether:

1. the instructions were clear
2. the questions were written in a way that was familiar to mathematics teachers (in a language that is used within their school routine, and ensuring that the linguistic complexity was appropriate to the target population, as suggested by Haladyna and Rodriguez (2013))

[10]Temporary teachers have the same responsibilities as permanent ones. Most of them work 40 hours being 25 hours of teaching and 15 hours of planning.

3. the questions were clear and unambiguous (whether the questions were asking what I intended to ask)

4. the teacher objected to answering any of the questions, and if so, why.

My main reason for doing this exercise was to help control measurement errors which could significantly influence the validity of the study (Schonlau et al., 2002), and bring essential inputs to the validity argument. PHASE 3 was conducted to guarantee that teachers were interpreting the answer options in the same way I intended which would allow me to make inferences about their choices. In addition, these reviews helped me to ensure that the questions did not only have a solid theoretical base, but were also relevant to classroom teachers.

According to Haladyna and Rodriguez (2013, p.106) this is also one of the most effective ways to make sure that all distractors are plausible. That is, "obtain or know what [respondents] will be thinking when the stem of the item is presented to them."

When PHASE 3 was conducted again (PHASE 3B), with the version in Appendix B.3, the same six teachers answered the MaTFAKi again and another two teachers participated for the first time:

Cindy has been teaching mathematics in FD state schools for nine years. She was teaching grade 6 and 7 (11 and 12 years-old students) when she participated in PHASE 3B. She has a teaching qualification in mathematics and a post-graduate certificate in 'Teaching Higher Education'. Her school is located in one of the administrative regions.

Adam has a total of 17 years' experience teaching primary and secondary levels. He has a teaching qualification in mathematics, a degree in Education and a post-graduate certificate in 'Management in Education'. He was teaching grades 8 and 9 (13 and 14 years-old students) when he participated in PHASE 3B. His school is located in the same administrative region as Cindy's.

Although a sample of eight teachers could be considered a very small sample, when the intention is to assess clarity of instructions or item wording, acceptability of formatting, or ease of administration, a sample of 10 or less is considered sufficient (Hertzog, 2008). A field test with a greater sample was conducted in a further stage (see Section 6.3).

In the following sections I will present the results of PHASE 3. In accordance with what I did with the researchers' participation, I will first present a full narrative of how I analysed one of the questions based on each teacher's answer. Subsequently, I will present examples of their answers and the changes made, according to the aims of this review.

6.2.1 A full narrative of how I analysed each question

When answered by the teachers, most of the initial ideas were corroborated. It is important to remember that *a*) differently from the researchers, they had a printed version, *b*) teachers were asked to choose only one option – the one they judged to be the best, *c*) the version presented to them already encompassed the researchers' suggestions (Appendix B.2), and *d*) I conducted PHASE 3 twice – first with six teachers and 13 questions and then with eight (same six + two) teachers and 24 questions. The full narrative presented here is in relation to Q1 (Appendix B.2).

While Cathy explained why she chose Q1.B she also explained her own approach to teaching inequalities to pre-empt and address this kind of mistake:

> In this first question, about the inequalities, I would choose B, the students solved like equations, because they did not consider the inequality. What I usually do, to prevent this, I show them with numbers. I write, for example, 3 > -1. Then, I show them that if they just multiply by minus 1, it's going to be -3 > 1 which is not true. So, they need to invert the sign as well. They see with numbers first, so when they are working with inequalities, they will better understand that the x now represents an interval and not just a number anymore.

When explaining, she mentioned the same idea discussed during the meeting with the researchers: the 'x' in inequalities represents an interval and that this is the reason for 'inverting' the sign when multiplying by minus 1.

Marjorie also chose Q1.B:

> Well, when we work with them on inequalities, this always happens. Because we have been working with equations for so long, they keep solving like equations. Those who are cleverer, they notice very fast that

> you should solve the same way, but observing the 'rule in relation to the inequality'. So here, they solved like equations.

In this case, it is possible to notice that she identified that students are solving like equations, but she mentions "rule in relation to the inequality". Thus, differently from Cathy, she focusses on 'the rule' instead of on the understanding of the conceptual differences between equations and inequalities. She then continued explaining her choice when comparing to Q1.C:

> *These students do not know how to solve inequalities because they did not find the correct results* [reading Q1.C], well... it's not that they don't know how to solve it, they are just so used to equations at this point that they do not take the inequality into consideration. And I don't think that we should say that they *don't know how to solve* [her emphasis] just because they didn't find the correct solution... You have to consider the steps they followed, the process. Yeah... option B is the best way of interpreting it.

Her comments regarding Q1.C confirmed the original idea behind the option – to consider that a student does not understand a concept based only on an incorrect final answer of an exercise.

For Amelia, both Q1.A and Q1.B were good, but according to her, she chose Q1.B because:

> Although they did make a mistake at the end, option B is the one that really explains that the students' mistake was that they were solving the inequalities as equations. That's what you're asking here, isn't it? The interpretation. Usually, we change the sign [pointing to the >] at the end, but in this case, because the students solved as equations, they didn't take the inequality into consideration – the arithmetical computations they know.

Her comment shows that she was being careful about what the question was asking and how she should judge the options and make her choice ("That's what you're asking here, isn't it?"). Her choice also corroborated the elements raised during the meeting with the researchers: Q1.A was just recognising the mistake (at the end), whereas Q1.B was an actual interpretation of the mistake.

Max, Jordan and Martin also chose Q1.B. Their explanations, however, did not include elements of any other option. Max, for example, said:

> This is something that I live with every day. It takes a while for students to learn equations. Especially when we have a negative coefficient with the 'x'. They understand that they have to multiply by minus 1. But when we move to inequalities, they don't understand that the 'greater than' sign needs to turn into a 'less than' sign as well. So, in this case it's clear that they are solving as equations.

Jordan commented: "Well, in this one I could say that they solved as equations, but did not understand what an inequality is, which is something that I see very often in my lessons, they got confused here."

Martin also brought up the idea of emphasising to students that they need to invert the sign when they multiply by minus 1.

> This a common mistake. Normally this is what happens when I teach inequalities. If we are not careful and show students, and draw their attention to the 'greater than sign', especially when the 'x' coefficient is negative... I always emphasise that; and I also explain that they don't need to bring the 'x' to the left side, they can keep it on the right, if they find it convenient.

The second time around, when they answered the version in Appendix B.3, all teachers kept the same answer. Cindy and Adam, on the other hand, who were answering for the first time, chose Q1.A. Adam explained:

> I like both A and B, but I think we need to be pragmatic. Although I know they were solving as equations, it is ok to solve as equations. In this case, at the end they forgot to flip the sign. And we know that this is a common mistake and it takes a while for students to get used to flipping the sign when they are dealing with inequalities. Yeah... I'll go with A.

Cindy's explanation followed the same line of argument:

> It is hard to choose here. Both A and B explain what happened. The students were solving as equations, and usually they forget to flip the sign at the end. I always draw their attention to that, but it takes a while for them, and I don't know why. I just keep reinforcing it until they get it. I'm going to choose A then.

Therefore, in both cases, they identified what was happening, but preferred to choose the option which had a more ‘pragmatic explanation’.

In general, I could also observe that when they said “this always happens”, “something that I live with every day”, “This is a common mistake” or “something that I see very often in my lessons” it demonstrates the relevance of the question as it evidently reflects issues they face in their lessons.

In conclusion, although at this stage there was not a substantial difference among teachers’ answers, from their comments, it was possible to conclude that they were interpreting the options as expected. Based on these results, I decided to keep this version of Q1 until the field test, to assess how it would behave when answered by a larger sample.

The summary of teachers’ answers from PHASES 3A and 3B can be found in Tables 6.7 and 6.8 respectively. The correspondence of the questions from all versions of the MaTFAKi can be found in Table 11.2[11].

Table 6.7: Teachers’ answers - PHASE 3A.

Question	Marjorie	Jordan	Cathy	Max	Martin	Amelia
1	B	B	B	B	B	B
2	B	B	B	C	D	A
3	A	A	D	A	A	A
4	B	C	A	B	C	D
5	C	A	C	C	C	C
6	B	C	B	B	B	C
7	B	B	D	A	A	B
8	D	C	A	D	C	C
9	D	D	D	D	D	D
10	D	C	D	C	D	D
11	D	B	C	A		A
12	D	A	A	D	A	D
13	C	A	B	A	B	A

[11] In some questions, the order of the response options is also different.

Table 6.8: Teachers' answers - PHASE 3B

Question	Marjorie	Jordan	Cathy	Max	Martin	Amelia	Adam	Cindy
1	B	B	B	B	B	B	A	A
2	B	B	B	C	D	A	B	D
3	A	D	D	A	D	A	A	A
4	B	C	A	B	C	D	B	C
5	D	D	B	B	D	B	B	D
6	C	C	C	C	A	C	A	B
7	C	A	B	C	A	A	A	C
8	D	D	D	B	D	D	D	B
9	B	C	B	B	B	C	C	B
10	B	B	D	A	A	B	D	B
11	D	C	A	D	C	C	A	D
12	D	C	C	C	C	D	D	C
13	C	D	C	A	B	A	A	C
14*	B	B	B	B	B	B	B	C
15	D	C	D	C	D	D	C	C
16	B	D	D	A	B	C	D	C
17*	C	D	A	A	B	B	B	B
18	B	B	B	A	B	B	A	A
19	A	D	A	B	C	D	B	A
20	C	B	C	B	C	C	B	B
21	D	A	A	D	A	D	A	A
22	C	A	B	A	B	A	A	C
23	B	B	B	B	B	D	C	B
24	D	C	A	D	A	B	D	D
*The order of the answer options is different from the equivalent questions in the previous version. The previous version is in Appendix B.2 and the version presented in this table is in Appendix B.3								

As explained, this thorough analysis was done with every question in every stage. In addition, I was also focussing on the other goals of this review. Therefore, in the next sections I show examples of teachers' comments and the changes I made according to the goals specified on page 125.

6.2.2 Checking whether the instructions were clear

During all the time that teachers were answering the MaTFAKi, I asked additional questions to check if they clearly understood what they were required to do.

I always started by asking if they understood how the MaTFAKi would work, whether they understood that they had to answer based on what they thought 'would be the best thing to do' and not what they 'would actually do' in their lessons. Jordan's answer to Q15 (Appendix B.3) exemplifies how they understood it:

> I believe that option C would be the best thing to do, but I also think that in our reality, with classes with so many students, I would do D, because in option C, it's very open. If you open this much, the students will make a mess[12], it would be difficult to bring their focus back to the task. However, as you asked me to answer what I think would be the best thing to do, I would choose option C, because I would make him think, it's a brilliant way of inciting their learning and to promote a great discussion. Thinking about the practical side, I would go for D.

In some cases, however, I noticed that some instructions needed to be rephrased, as was the case with the transition between Q9 and Q10 (Appendix B.3).

I realised that, as Q10 depended on the answer of Q9, if the respondent chose the same feedback as *Mrs Brown*, there was no problem and they would understand Q10 and present a reasonable explanation for their choices.

However, when the feedback they chose as the best one was different from *Mrs Brown*'s, they would not switch their minds to *Mrs Brown*'s feedback; the rationale of their answers would be based on the feedback they chose in the previous question, and that was not the intention. Therefore, I changed the stem of the next question to draw their attention to *Mrs Brown*'s feedback (see Table 6.9).

I also felt that it was necessary to verify if they understood the ideas when the scenario was long and involved many aspects and interpretations. In this case, I also asked them if they understood how the ideas were flowing along the questions. So I asked questions such as: Can you explain in your own words what is happening? Can

[12]Here, the teacher said that students would "fazer um auê" which is a common expression in the Brazilian Portuguese informal language. Although I translated as 'make a mess', 'fazer um auê' is not always seen as a bad behaviour, it could mean, for example, that the students would be very excited and would like to participate in a non-organised way. It gives a more joyful connotation. The best translation of 'make a mess' would be 'fazer bagunça' which is not what the teacher intended to say in this case.

Table 6.9: Different versions of Q10 stem.

Previous version	Revised version
What would be the best thing for Mrs Brown to do in the next lesson, after giving written feedback to these students?	Based on the feedback that Mrs Brown wrote, what would be the best thing for her to do in the next lesson?

you recall the order of the events? These kinds of questions also provided valuable information for writing items in a way familiar to teachers as I explain below.

6.2.3 Analysing whether the questions were written in a way familiar to teachers

Asking for teachers' rationale for choosing one option over another, allowed me to check whether the questions were written in a way familiar to them, because they would explain it in their own words, which meant I could compare it to what was written.

In some other cases, I specifically asked some questions to check their familiarity with the language. That was the case, with *criteria* in Q21 (Appendix B.3). When validating this version of the MaTFAKi with the researchers, some of them raised a concern in relation to whether the teachers would understand what I meant by *establish some criteria*. Therefore, before showing Q21 to the teachers, I always asked what they thought that I meant when I asked about *establishing some criteria*.

Even if all eight of them interpreted it in the expected way, to guarantee that all teachers would do so, I decided to include *that the question should meet* at the end of the sentence to make it even clearer (Table 6.10).

Table 6.10: Different versions of the instruction before Q21.

Previous version	Revised version
However, for students be able to do the activity it was necessary to establish some criteria.	However, for students be able to do the activity it was necessary to establish some criteria that the questions should meet.

The same was done with *peer-assessment* in Scenario 4 (Appendix B.3). My previous experience with teacher training or even talking to teachers, has shown that sometimes, Brazilian teachers tend to interpret peer-assessment as one student helping the other during a test (which is a common practice in Brazil as a way of 'varying' the way teachers assess their students). However, as the term peer-assessment actually "involves students in assessing each other's work, through reflection on the goals and what it means to achieve them" (Sebba et al., 2008, p. 6), I felt that I should include additional words to aid clarity, and that it was also necessary to check whether they understood the activity as an example of peer-assessment.

Therefore, in the Portuguese version, I included a small explanation in parenthesis (um aluno avaliando o outro)[13]. See Table 6.11.

Table 6.11: Different versions of the beginning of Scenario 4.

Previous version	Revised version
[...] estratégia de avaliação entre pares para avaliar se seus alunos aprenderam a identificar e aplicar as definições e propriedades de raio e/ou diâmetro para calcular comprimento e área.	[...] estratégia de avaliação entre pares *(um aluno avaliando o outro)* para avaliar se seus alunos aprenderam a identificar e aplicar as definições e propriedades de raio e/ou diâmetro para calcular comprimento e área.

6.2.4 Changing those questions that were not asking what I intended to ask

As PHASE 3 was conducted face-to-face, I could ask teachers to explain their reasons for choosing one option over the others. This exercise provided very important information regarding how they were interpreting the questions and the response options.

For most questions, teachers confirmed my initial intentions and also the suitability of the changes the researchers suggested. In some other cases, changes were still necessary because different explanations were being given. That was the case, for example, with Q3 (Appendix B.2) or Q10, 17 and 18 (Appendix C.2).

Apart from Cathy, all teachers chose Q3.A. However, their explanations were different. Cathy said:

> I don't like to tell the students that they got it wrong. First because it usually puts them off. Second because they did not get everything wrong, they just did not pay attention to the greater than sign. So, straight way, I wouldn't

[13]In English, 'one student assessing the other'.

choose either B or C. For me, option D is the best. I prefer using different examples and then students go back to the ones they solved before and try to identify their mistake. So, I would draw their attention to the sign using numbers and different inequalities and then let them go back to the same inequalities and fix them.

Cathy's comments showed she was interpreting Q3.D as expected – the teacher was focussing on the learning intention rather than on the activity itself.

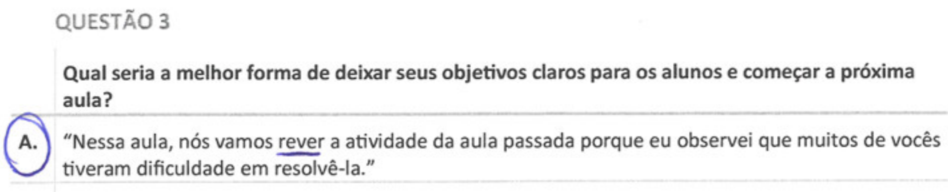
However, the same could not be said in relation to other teachers' answers. Marjorie said: "I always do this – tell students what I will do and why. So, in this case, I would do A: go over the activity from last lesson because I noticed that many of them had trouble doing it."

When questioned about the other options, Marjorie explained: "I think it makes more sense to them if we draw their attention to those inequalities that they just did. It's still fresh in their minds."

Amelia, at the beginning, emphasised her comments in another part of Q3.A: "I would choose option A because it says that I observed that a bunch of students had problems with it."

When she rephrased the sentence using "a bunch"[14], I understood that she was choosing Q3.A because it said that the teacher noticed that *many* students had had problems. To confirm, I asked why, for example, she preferred Q3.A to Q3.D.

She answered: "Because in option A it says that we are going *to review* [her emphasis], which means that students will do an activity they've seen before, and in D, it's more general, it says that we would solve some inequalities." She even underlined the word *review* (rever, in Portuguese) to show which word drew her attention and made her choose Q3.A (See Figure 6.1).


QUESTÃO 3

Qual seria a melhor forma de deixar seus objetivos claros para os alunos e começar a próxima aula?

A. "Nessa aula, nós vamos rever a atividade da aula passada porque eu observei que muitos de vocês tiveram dificuldade em resolvê-la."

Figure 6.1: Amelia's answer to Q3.

[14]The words *a bunch* were chosen as the translation of *um monte* which is an informal way of saying *a good amount of, many, the majority* in Portuguese.

Therefore, she was not choosing Q3.A for the reason I originally thought. She actually judged the same as Marjorie – that the idea of Q3.A was that the teacher would go through the *same inequalities* and in Q3.D, the teacher would bring *other random examples* instead of approaching the solution of inequalities in general.

Max also brought up elements of other options:

> In this case, I prefer option A because it says *review*, and not *correct*, like it says in option C, or *I* will show, like option B. I don't like to say that *I will correct*, because those students who got it wrong will think: 'Gosh, I got it wrong *again*' [his emphasis]. And the last one, it's too vague: *solving some inequalities*. If we review the ones they already solved, we can focus on their mistakes and show them all the important aspects, like the negative coefficients, the greater than and less than signs, and so on.

Jordan explained that Q3.A was the best, comparing with Q3.B first:

> [...] in option B it says categorically that they solved incorrectly. However, they didn't get everything wrong. They got it wrong at the end. So, I think this would be the best way to put it, say they had trouble instead of saying they got it wrong, like it says here [pointing to option B].

With this explanation, it was possible to confirm one of the intentions of Q3.B: the teacher recognising that it says *solved incorrectly*.

As he already mentioned another option, I also asked about Q3.D. He said: "Well... option D is actually pretty good as well, because it also says that students had trouble instead of saying they got it wrong. However, I think we need to review. To go over the whole concept, and not just bring new exercises."

Martin followed the same principle:

> Option A, because in this case, if I go over the whole activity, I'll probably be able to address other questions from students, review some concepts related to the solution of inequalities to help them understand why they're making those mistakes.

As can be seen, teachers gave different explanations for the same option and the same explanation for different options. Comparing Marjorie, Amelia and Max's explanations with Jordan and Martin's, it can be concluded that there was a problem with the way they were interpreting Q3.A and Q3.D.

When I included *solve some inequalities* in Q3.D, my intention was aligned to what Jordan mentioned at the end of his explanation – "go over the whole concept" – or to what Cathy said, "I would draw their attention [...] using different inequalities and then they go back to the same inequalities and fix them" or either of Martin's comments: "address other questions from students, [and] review some concepts related to the solution of inequalities [...]". However, as Martin and Jordan chose A and Cathy chose D (based on the same rationale), it showed me that something else needed to be changed.

The solution that I found (see Table 6.12) was to change *solving some inequalities* to *remembering how to solve* as this would give the idea of going over the concept, but focussing on the concept itself and not the activity.

Table 6.12: Different versions of Q3.D.

Previous version	Revised version
"We are going to start our lesson by solving some inequalities because I noticed that some students had trouble with the activity from the last lesson."	"We will begin our lesson remembering how to solve inequalities because I noticed that some students had trouble solving them in the last class activity."

When I did the review for the second time (with the version in Appendix B.3), I could confirm that this small change worked, as Jordan and Martin changed their answer to D, and kept the same explanation. Whereas Marjorie, Amelia and Max kept their initial answers explaining that they believed that it would be a better approach to focus on the same inequalities. Cathy also kept her initial answer with the same explanation.

Both Adam and Cindy followed Majorie, Amelia and Max, favouring Q3.A. For Adam "A is much better because I can show them that *their trouble* [his emphasis] was to forget to flip the sign. I like to draw their attention to their mistake."

Cindy's explanation also focussed on 'flipping the sign':

> I like both A and D, but I believe that in this case their mistake was very simple, I would just show them that they did everything right but forgot to flip the sign. I would do D if I noticed that they didn't understand anything, but I don't think it is the case here.

In both cases, it is possible to say that their arguments were aligned to their answers to Q1.

6.2.5 Identifying issues on the question that teachers objected to answering

The only question teachers objected to answering was Q11 (Appendix B.2). According to Martin:

> Every option is indicating that there is a problem with the exercises, and I don't think there is. I don't think it is not in accordance with the teacher's intention. I believe that, with these three exercises, he opens up a good opportunity for discussion with students in relation to division by half. I would choose one if you had an option like: *You're good to go. Your exercises are appropriate for your intention*. Because you don't have this option, I prefer not to choose any of them. [Martin, Q11, Appendix B.2]

This also happened with Amelia, who said: "I wouldn't give any of these as advice to *Mr Fitzgerald*, because even this one [referring to exercise II] which is not division, when you divide by half, you multiply by two, so it is a different way of working with division by half." Therefore, although she recognises that the activity does not require students to divide by half, she still thinks that they are appropriate for the teacher's intentions when she says that *it is a different way of working with division by half*.

However, because she thought she needed to select an option, she chose Q11.A because "I think you should review your assignment because the language in the exercises is a little complicated and the students might not understand that the teacher wants them to divide by half." However, as Q11.A did not have this content, and in line with the researchers and my own view it was the option considered to be the best answer, she was choosing the best answer even though she did not have the knowledge being assessed. This situation reinforced the need to review the options in this question.

When I initially developed Q11[15] I tried to follow Haladyna and Rodriguez's (2013) guideline regarding plausibility of the options. I believed that teachers would notice that somehow the exercises had some problems and therefore it was not plausible that a teacher would choose an option affirming that all exercises were in accordance to *Mr*

[15] See Table 11.2 for the other versions.

Fitzgerald's intention. Therefore, I decided not to include an option with this content. However, the results from PHASE 3A showed the opposite.

After the review with all teachers, I decided to exchange the options as shown in Table 6.13.

Table 6.13: Different versions of Q4.

Previous version	Revised version
I think you should review your assignment because none of your problems are assessing what you want to assess.	I think your activity is in accordance with your goal and therefore is ready to be given to students.

I decided to exchange this specific option because it also raised some concerns with Cathy's answer. She believed that it was not assessing division by half because "the students could be adding halves until they get the answer (2 cups and a half) and would not actually calculate that 2 cups and a half is $\frac{5}{2}$ and then inverting $\frac{1}{2}$ to multiply."

Therefore, I would not be able to affirm with certainty what was the knowledge that this teacher had – whether s/he was considering that knowing how to divide by $\frac{1}{2}$ is to know the invert-and-multiply rule, and therefore that was why s/he chose that option; or if it was because s/he identified any other problems with the options. In addition, I would not be able to differentiate among teachers as in both case their answers could be considered correct.

After consulting with the researchers again, I presented my solution to the problem (Table 6.13) and they confirmed that that was a good way to solve it. They all agreed that knowing how to divide by half involves understanding the concept and not just learning how to use the invert-and-multiply rule. It is correct to say that a student knows how to divide by half if they calculate how many $\frac{1}{2}$s can be fitted into $\frac{5}{2}$ as is the case with the problem included in the MaTFAKi.

In PHASE 3B, Martin and Amelia confirmed their views and chose Q17.B[16]. Cathy chose Q17.A but still commented: "I will choose option A because it is more aligned to what I think, but I don't think we should say *some*, because none of them are."

The questions aforementioned are just examples of some of the many changes that I made based on the review with the teachers (the different versions can be found in the appendices). The main purpose was to provide an idea of the systematic and reflective

[16]The order of the options is also different. Option B here replaced option C in the previous version.

approach that I followed to produce the MaTFAKi. Once I had a fully revised version with 24 questions, a field test was conducted.

6.3 PHASE 4: Field test

After spending over a year going through all these phases with the researchers and teachers in Brazil, I managed to have a 24-question MaTFAKi – four questions for each DoK. The next stage, was to conduct PHASE 4: a field test with a larger sample of mathematics teachers. The main purpose of PHASE 4 was to assess the psychometric properties of the MaTFAKi and decide which questions would comprise the final version.

6.3.1 Sampling and delivery considerations

Although my ultimate target population was FD mathematics teachers, I conducted PHASE 4 with teachers from outside the FD. The main reason for that was because my target population was not large enough for it to be possible to use some of them for this field test. Usually with online surveys, the response rate tends to be small, and therefore I could not risk not having enough teachers for the main survey.

To analyse how valid it would be to use the results of PHASE 4 in evaluating the MaTFAKi, I included questions to collect background information so I could later compare the characteristics of the test sample with the final sample. I also made sure to use a similar approach when choosing the sample and delivering the MaTFAKi.

The MaTFAKi was sent to the schools' email with a request for the principal to forward the email to the mathematics teachers. Due to this two-stage process, I expected to have a low response rate. As Brazil is a huge country with thousands of schools, I decided to send many more emails than I would need if I was following a single-stage sampling strategy.

The email addresses and/or telephone numbers of all schools were retrieved from an official government website[17], which provides contact information for all schools in Brazil and information about their characteristics. After collecting the email addresses, a list of 5000 was randomly generated using *R*. This list was generated separately by state to maintain the proportion of schools. From the 5000 emails sent, 778 (15.58%) bounced back, which resulted in 4222 emails successfully sent.

[17] http://www.dataescolabrasil.inep.gov.br/dataEscolaBrasil/.

In the covering letter, I made it clear to respondents that the MaTFAKi was long, but they could close it and open it again to finish answering at a time suitable for them. I also included a question asking if they wanted to participate in further research so I could invite them for the retest (see Section 6.3.2) or talk to them about specific items that did not behave the way I expected and whether the instructions were clear. Although I included this final question, I decided not to tell teachers that this was a 'test', to avoid drop outs and to encourage them to take it seriously.

The pilot version was available for a month. As explained in my email, after two weeks I sent a reminder to the principals. By the end, I had 222 answers which represented 5.26% of the delivered emails. However, this is just an estimate as it is not possible to determine how many emails were read, how many were actually forwarded to teachers, and to how many teachers. An email thanking teachers and schools for their participation was sent at the end of PHASE 4. For those teachers who provided their email addresses, the 'thank you' message was sent to them directly.

6.3.1.1 Characteristics of the field test respondents

The sample for the field test included mathematics teachers from independent and state schools in Brazil. The teachers participated in the field test with no external rewards or motivational offers.

The average age of the respondents was 40 years-old, the youngest being 22 and the oldest 64. 51% of the respondents were female. Most respondents (88%) had a teaching qualification in mathematics and were teachers with less than 15 years of experience (55%). The majority (89%)[18] teach in state schools. There were no respondents from two states, both in the North of the country.

6.3.2 The results of the statistical analysis

My main goal with PHASE 4 was to gain an indication of the most robust items that I would later incorporate into the final MaTFAKi. For that, I carried out reliability and factor analyses. For all statistical analysis I used the free software environment for statistical computing and graphics *R*[19].

[18] 89% out of 181 teachers because 40 respondents were not teaching at the time they answered the MaTFAKi – acting as head teachers, on study leave, etc.

[19] More information at https://www.r-project.org/.

6.3.2.1 The results of the factor analysis

Exploratory Factor Analysis (EFA) is a statistical technique that can be used for exploring variable structures in datasets (Field, Miles, & Field, 2012). EFA techniques are based on the identification of clusters of highly correlated variables. These clusters can be interpreted as underlying dimensions in the data, and called factors or latent variables. EFA enables researchers to reduce the data to a small number of explainable concepts while accounting for the maximum amount of common variance. In this study, factor analysis was conducted based on the polychoric correlation matrix, which is more suitable for ordinal variables than the Pearson covariance matrix, intended for continuous variables (Carroll, 1961).

However, before running the factor analysis, I first checked if the data was appropriate. Three tests were run:

The Bartlett's test was highly significant ($\chi^2_{276} = 828.18, p < .001$). This means that the R-matrix was not an identity matrix and therefore there were some relationships between the variables, making factor analysis appropriate.

The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis (Overall KMO = .54). Although considered as 'mediocre' by Hutchenson and Sofroniou (1999), a value above .5 suggests that a factor analysis of the variables was appropriate (Field et al., 2012; Kaiser, 1970).

The determinant of the correlation matrix was calculated to detect multicollinearity, and resulted 0.02017236, which is greater than the necessary value of 0.00001 and therefore did not seem problematic.

Based on these results, I concluded that EFA could be conducted. The first step was to obtain the eigenvalues for each component in the data. With the factor analysis, 5 factors were over Kaiser's criterion of 1. For the Principal Component Analysis (PCA), 10 components had eigenvalues over 1.

The scree plot was slightly ambiguous and showed inflexions that would justify retaining 2 or 5 factors if factor analysis was conducted and 2, 5 or 10 for a PCA (See Figure 6.2).

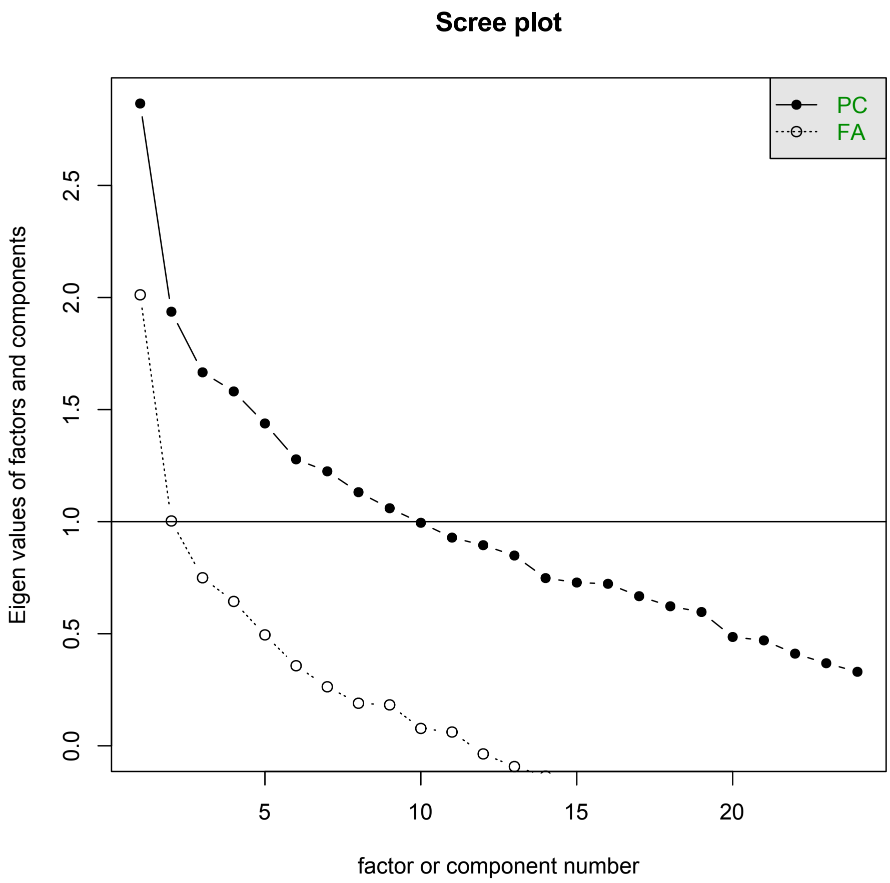


Figure 6.2: Scree plot for factors and components.

All those options were run and analysed. However, as reported in other studies using SJTs (Chan & Schmitt, 1997; McDaniel & Whetzel, 2005), the analyses did not provide interpretable factors, but confirmed the multidimensional nature of the MaTFAKi. As Chan & Schmitt (2002) argued, the difficulty that researchers face in obtaining interpretable latent factors of SJTs is probably due to the multidimensionality inherent in the situations and solutions.

6.3.2.2 The results of the reliability analysis

Reliability can be defined as the degree to which "measurements of individuals on different occasions, or by different observers, or by similar or parallel tests, produce the same or similar results" (Streiner, Norman, & Cairney, 2014, p. 9). That is, "Reliability is concerned with the stability and consistency of the actual measuring instrument or procedure" (Lyons & Doueck, 2010, p. 127).

There are different ways of analysing the reliability (test-retest, split-half, internal consistency, etc.) of an instrument, which mainly depends on the nature of the data. Usually, to assess reliability of knowledge questions, test-retest or split-half are used (Radhakrishna, 2007; Rust & Golombok, 2009).

The idea of the test-retest measure is to examine the stability of the scores over time (Multon, 2010; Rust & Golombok, 2009) by correlating the scores of the same respondent, under the same circumstances, with an interval between administrations. In a way, it consists of triangulating the same method at different times. On the other hand, the split-half measure is appropriate when one has many items measuring the same thing. However, it is not so appropriate when you have an instrument measuring many different things with different questions, and therefore it is difficult to have a proper half (Anderson, Gerbing, & Hunter, 1987; Streiner, 2003). Internal consistency is usually calculated through coefficient alpha and is more suitable for unidimensional instruments (Anderson et al., 1987; Henson, 2001; Streiner, 2003).

The ordinal alpha of the 24-question MaTFAKi used in the field test of this study was $\alpha = .67$[20], which can be considered acceptable for a new instrument (Briggs & Alonzo, 2012; Rust & Golombok, 2009) and compared to other SJTs used in previous research (e.g. Chan & Schmitt, 2002; Pulakos & Schmitt, 1996). However, this estimate may not be a meaningful index of internal consistency, given that item responses are likely multidimensional or if we view FA in mathematics as an aggregate composite ability consisting of multiple unitary or multidimensional DoK (both regarding FA and mathematics teaching and learning). That is, the scale and item heterogeneity makes alpha an inappropriate reliability index.

For this reason, researchers argue that test-retest or parallel forms reliability may be a more accurate approach to evaluating the reliability of any given SJT (Cabrera & Nguyen, 2001; Christian, Edwards, & Bradley, 2010).

Therefore, based on the multidimensionality of the MaTFAKi and because the main reason for this field test was to select appropriate items for the final version, I decided on the test-retest measure. It provided me with information of not only how 'stable' the MaTFAKi was, but also how *each item* was behaving.

Empirical research shows a range of test-retest reliabilities reported for different

[20] Although Cronbach's α is the most used coefficient, in this study I used *ordinal alpha* (Gadermann, Guhn, & Zumbo, 2012; Zumbo, Gadermann, & Zeisser, 2007), which more accurately estimates alpha for measurements involving ordinal data (Garrido, Abad, & Ponsoda, 2013; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010).

SJTs. Ployhart and Ehrhart (2003), for example, report test-retest reliabilities ranging from as low as .20 to as high as .92. In this study, Table 6.14 shows the range from .50 (Q11) to .97 (Q17).

Table 6.14: Test-retest reliability.

Question	Test-retest reliability
1	.95
2	.76
3	.57
4	.84
5	.85
6	.92
7	.79
8	.79
9	.85
10	.86
11	.50
12	.54
13	.85
14	.54
15	.81
16	.85
17	.97
18	.89
19	.85
20	.71
21	.82
22	.91
23	.85
24	.94
$n = 53$	

Based on the results of the test-retest reliability, I decided to exclude Q3, Q11, Q12 and Q14 because these were the ones in which $r \leq .70$ (.57, .50, .54 and .54 respectively). Observing Table 6.6 on page 123, Q11, Q12 and Q14 had also caused some division in the researchers' opinions with regards to the alignment between those questions and the DoK that they were developed to measure.

After choosing the questions that I would keep for the final version, I conducted the reliability analysis again, which yielded $\alpha = .64$.

Therefore, after excluding these four questions I ended up with a 20-question MaTFAKi that could be considered a reliable instrument for measuring different aspects of FA in mathematics teaching and learning in Brazil.

6.4 Summary

In this chapter, I have explained and reflected on the phases that I followed after writing the questions that would make up the MaTFAKi.

In addition to carefully developing the questions myself, a small-scale pilot study was conducted to analyse researchers' then teachers' reactions to the MaTFAKi and to assess the validity of the feedback information that I would be able to provide with the results of the study. A larger scale field test was conducted to assess the psychometric properties of the MaTFAKi. Figure 6.3 on the next page shows the steps followed and the goals achieved during all these phases.

All these steps were conducted to make sure that the MaTFAKi was not only valid and reliable, but formatively useful. In summary, the quantitative data provided numerical descriptions of the reliability and validity of a new instrument, while the qualitative data assisted in its content and construct validation. In the next chapter, I present the final version of the MaTFAKi, with the rationale of each question.

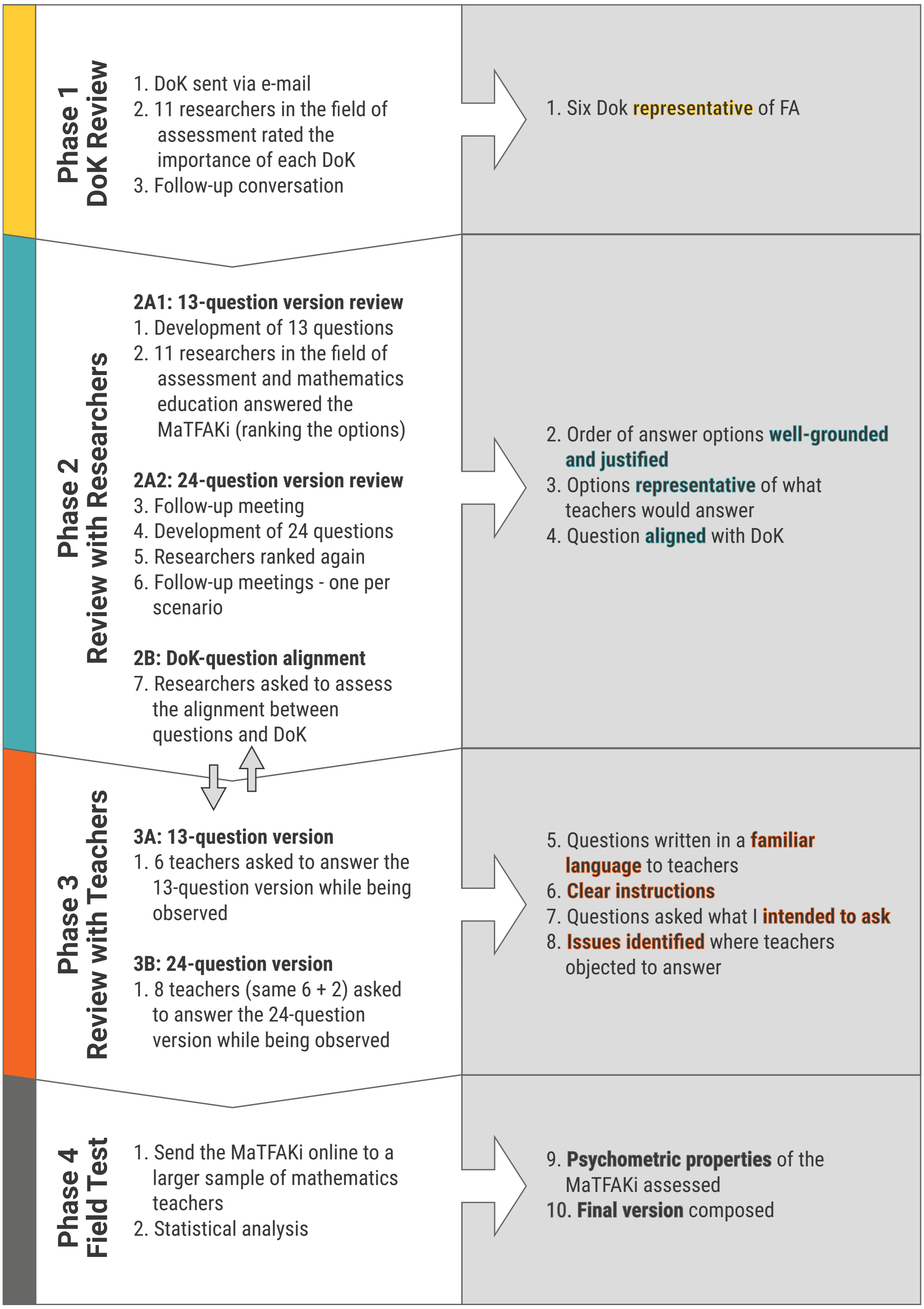


Figure 6.3: Pilot phases and what was achieved in each of them.

CHAPTER 7

PRESENTING THE MATFAKI

This chapter presents and explains the MaTFAKi in detail. This explanation comprises not only the rationale for including each question, but also the ranking order of the answer options.

Each section represents a Scenario from the MaTFAKi.

7.1 Scenario 1: Feedback to whole group

Scenario 1 has already been explained in Section 5.4.1. Tables with the scores and explanations of each question were also included in that section.

The choice of starting the MaTFAKi with a short scenario was deliberate, to not 'scare' the teachers, and also to get them used to how the MaTFAKi would work. Scenario 2 is relatively more elaborate and requires more reading.

7.2 Scenario 2: The teacher walking around while students work in groups

Scenario 2, which was developed from one of the questions of Ball et al.'s study (2008, p. 400), was originally Scenario 5. However, as it required a lot of reading and interpretation, I decided to move it because, during PHASE 3, I noticed that teachers were tired towards the end of the MaTFAKi and this was causing them to not pay the necessary attention to this relatively elaborate scenario.

In this scenario, teachers are required to analyse whether an activity is in accordance with the learning intentions set out at the beginning (Q4). After that, they have the opportunity to interpret one student's mathematical thinking based on a conversation that he is engaging in with other students (Q5). As this conversation elicits a mathematical misconception and some confusion with the learning intentions written on the board, teachers have to identify what could have been done to avoid this confusion (Q6); and to judge how the teacher could help students to use the information they have to reach an agreed conclusion (Q7).

The scenario, therefore, starts with *Mr Fitzgerald*'s intentions and some problems he prepared (Figure 7.1).

Scenario 2: 4 questions

Mr Fitzgerald has been teaching his 7th year students how to divide by fractions. For his next lesson, he has decided to assess whether his students know how to divide by ½. For that, he has developed three problems:

1. **Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?**
2. **You have £1.25 and may soon double your money. How much will you end up with?**
3. **Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?**

Before assigning them to his students, he decided to show you the problems and ask whether they are appropriate for his intention or not.

Figure 7.1: *Mr Fitzgerald*'s activity.

Q4 (Figure 7.2) is introduced in the same page. Teachers are required to compare the three problems with *Mr Fitzgerald*'s intentions and assess whether the activity is appropriate or not (an example of KAM).

Question 4. What do you think is the best advice to give to Mr Fitzgerald?

- "I think you should review your assignment because some of your problems are not assessing what you want to assess."
- "I think you should include exercises which require students to divide by fractions other than ½."
- "I think your activity is in accordance with your goal and therefore is ready to be given to students."
- "I think you should include more questions because three are not sufficient for getting the information that you are looking for."

Figure 7.2: Question 4.

The explanation of each answer option, *in order of adequacy*, is presented in Table 7.1.

Table 7.1: Question 4 scores and explanations

Option	Score	Explanation
A	4	Teachers identify that the activity needs revision. The reason is because some problems are not assessing what *Mr Fitzgerald* wants to assess (division by $\frac{1}{2}$). Teachers consider *Mr Fitzgerald*'s intention.
B	3	Teachers identify that the activity needs revision. The reason is because they feel it should include exercises that require students to divide by fractions other than $\frac{1}{2}$. Teachers do not consider *Mr Fitzgerald*'s intention.
D	2	Teachers identify that the activity needs revision. The reason is because three problems are not enough for getting the information the teacher is looking for. Teachers do not consider *Mr Fitzgerald*'s intention.
C	1	Teachers do not identify that the activity needs revision and therefore no reason is given. Teachers do not consider *Mr Fitzgerald*'s intention.

According to Siegel and Wissehr (2011), being able to identify that exercises are not in accordance with the learning intention is a key ability to be considered assessment literate. In addition, if classroom assessment information is of poor quality or incomplete, a teacher will not be able to effectively interpret and communicate information about students' learning (Brookhart, 1998).

Q4.A was considered the best option because it is the only one where teachers identify that the activity is not in accordance with *Mr Fitzgerald*'s intention, as problem III is the only one which requires to divide by half, and therefore will not provide credible information in regards to students' learning (Earl & Katz, 2006).

In Q4.B and Q4.D, teachers also identify that there was a problem with the activity. However, even though both include plausible explanations, teachers do not consider *Mr Fitzgerald*'s intention.

In Q4.B, even if the teacher includes exercises that *require students to divide by fractions other than* $\frac{1}{2}$, it will not provide the information s/he is looking for, affecting on how effectively s/he will interpret and communicate students' learning (Brookhart, 1998). Equally, in Q4.D, the inclusion of more exercises will also not solve the problem as some exercises will still not assess what *Mr Fitzgerald* wants to assess.

Q4.B was considered to be better than Q4.D because the suggestion is to assess other learning intentions, whereas in Q4.D the teacher is only concerned with the amount of exercises.

Q4.C was considered to be the least appropriate because teachers do not recognise any problems at all. Even if it could be argued that in problem II, doubling is the same as dividing by half, problem I involves dividing by two and therefore it is clear that the problem is not in accordance with *Mr Fitzgerald*'s intention. Again, the task will not provide good information for the teacher regarding students' learning (Black et al., 2003).

The scenario then moves to *Mr Fitzgerald* writing his learning intentions on the board (Figure 7.3).

At the beginning of the next lesson, Mr Fitzgerald felt it was important to make clear to students what the learning intentions were and wrote on the board:

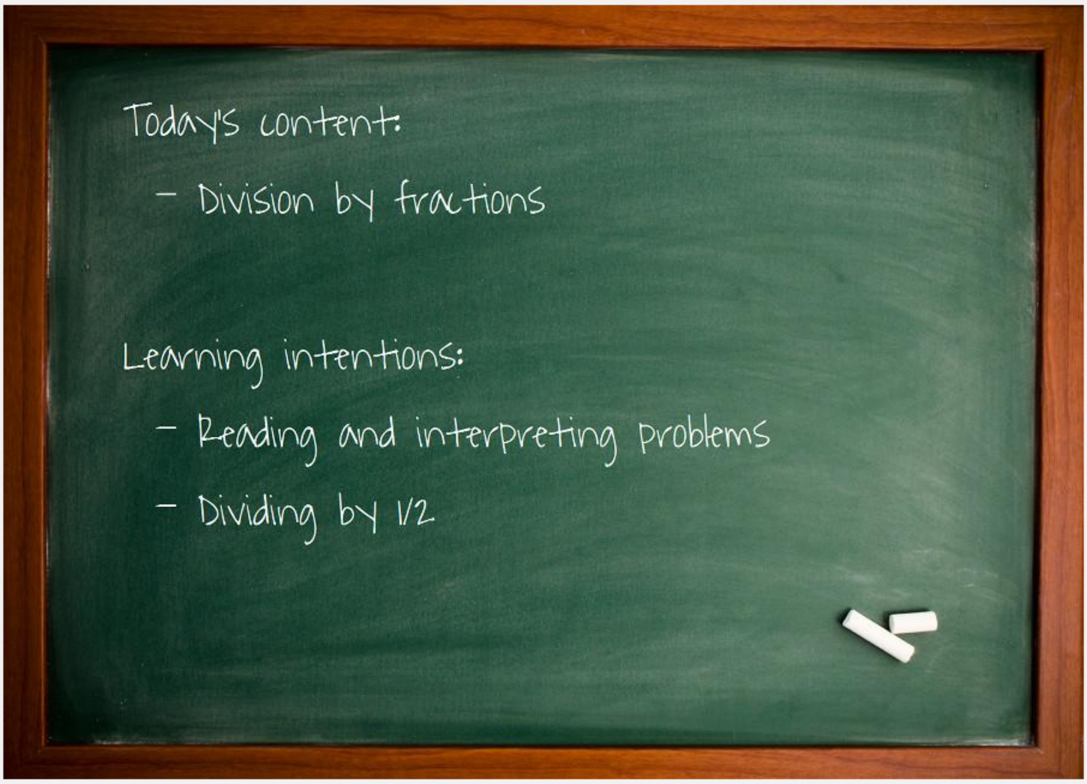


Figure 7.3: Board with *Mr Fitzgerald*'s learning intentions.

On the next page, teachers are introduced to a short explanation and a conversation among three students (Figure 7.4), before being introduced to Q5 (Figure 7.5).

After that, Mr Fitzgerald gave the activity to students and put them into trios so they could solve the exercises together.

Meanwhile, Mr Fitzgerald walked around the room answering questions and observing students. When passing by the trio Daniel, Carl and George, he heard the following conversation:

Daniel: Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

George: We just need to do 1 + ½ and divide by ½.

Daniel: Yes.

Carl: No, guys. She wants to divide it between 2 friends. So we have to do 1 + ½ and divide by 2.

Daniel: But it is written on the board that the exercises are to divide by half. So we have to divide by half.

Carl: But it is also written that we need to read and interpret the problems, so I think it's a tricky one.

George: Ok... but even though... if she wants to divide into two friends, she wants to divide by half, which is what I have said since the beginning.

Figure 7.4: Conversation among *Carl*, *George* and *Daniel*.

Question 5. Based on this conversation between the students, what can Mr Fitzgerald conclude about George's learning?

- George is dividing by half because he is having trouble interpreting the problem.
- George is considering divide by half and divide in half as the same thing.
- George is dividing by half because of the learning intentions written on the board.
- As the problem did not address division by half, little can be said about George's learning.

Figure 7.5: Question 5.

Differently from Scenario 1 in which the interpretation of students' learning is done through students' written answers, in Q5 the interpretation is done through analysing their talk (another example of KIE). Table 7.2 shows the scores and explanations of each option.

Table 7.2: Question 5 scores and explanations

Option	Score	Explanation
B	4	Teachers' interpretation is that *George* is considering divide by half and divide in half as the same thing. Teachers consider the conversation as a whole, and identify a mathematical misconception in *George*'s learning.
A	3	Teachers' interpretation is that *George* is having trouble interpreting the problem, which could be considered a plausible interpretation. However, teachers do not consider the conversation as a whole, and do not identify a mathematical misconception in *George*'s learning.
C	2	Teachers' interpretation is that *George* is being influenced by the learning intentions written on the board, which could be considered a plausible interpretation. However, teachers do not consider the conversation as a whole, and do not identify a mathematical misconception in *George*'s learning.
D	1	There is no interpretation. Teachers do not consider the conversation as a whole, and do not identify a mathematical misconception in *George*'s learning.

Q5 is comprised of one option (Q5.B) in which the teacher's interpretation takes the whole conversation into consideration, and therefore arrives at the best conclusion that *George* is considering divide by half and divide in half as the same thing; and one (Q5.D) that does not take the conversation into consideration at all and does not provide any interpretation about *George*'s learning. The other two options (Q5.A and Q5.C) present two different interpretations.

In regards to Q5.A, even though research has shown that some troubles that students have with solving word problems often comes from the difficulty in interpreting the problem (e.g. Bernardo, 1999; Hegarty, Mayer, & Monk, 1995; Martiniello, 2008), this cannot be said in relation to the situation presented. This could be argued if *George* had not said the last part when he insisted *If she wants to divide into two friends, she wants to divide by half....* That is, he clearly shows that he understood that the problem says *divide into two friends*, but then he states that it is the same as divide by half, even after *Carl* has disagreed with him and explained his rationale.

Equally, in Q5.C, the interpretation that the learning intentions are the problem is not possible because during the conversation *Carl* mentions the correct solution and also draw the groups' attention to the other learning intention about interpreting the problem. Again, when analysing the whole conversation, there are enough elements to conclude that *George* was considering diving by half and divide in half as being the same thing.

Q5.A is considered to be better than Q5.C because it provides an interpretation that indeed can be a learning gap, whereas Q5.C 'blames' the learning intentions written on the board, and therefore offers an interpretation that is not a gap in *George*'s learning.

The scenario continues with a short text between Q5 and Q6 (Figure 7.6).

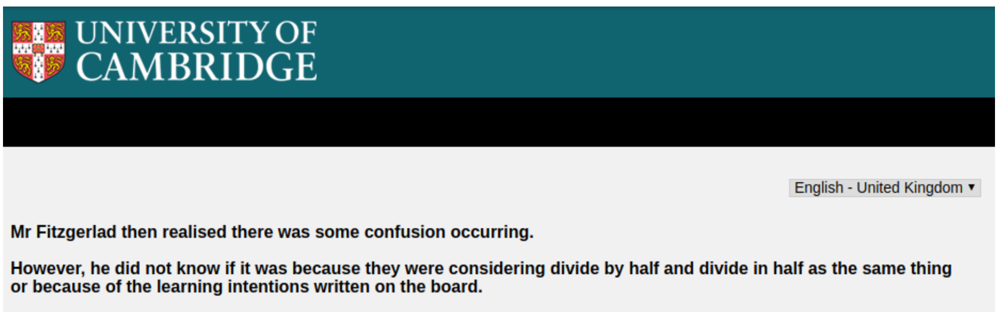


Figure 7.6: Text between Q5 and Q6.

Q6 (Figure 7.7) is introduced to address the issue with the learning intention raised in Q5 (an example of KLI).

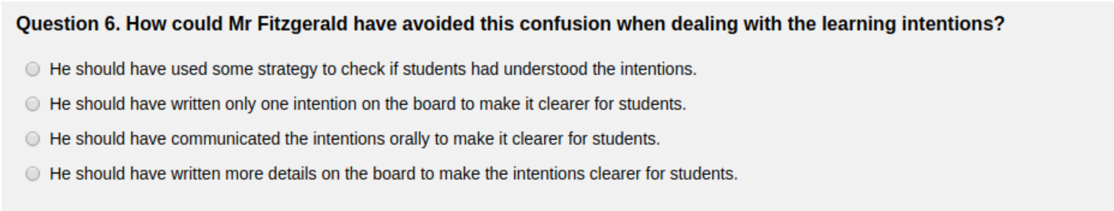


Figure 7.7: Question 6.

Q6 requires teachers to show whether they recognise the importance of making the learning intentions clear to students, and of checking whether students understood them (Wiliam, 2011). Table 7.3 shows the scores and explanations of each option.

Table 7.3: Question 6 scores and explanations

Option	Score	Explanation
A	4	Teachers identify that it is important to use some strategy to check if students understood the learning intentions written on the board.
D	3	Teachers do not identify that it is important to use some strategy to check if students understood the learning intentions. For them, writing more details on the board would make it clearer for students.
B	2	Teachers do not identify that it is important to use some strategy to check if students understood the learning intentions. For them, writing only one intention on the board would make it clearer for students.
C	1	Teachers do not identify that it is important to use some strategy to check if students understood the learning intentions. For them, communicating the learning intentions, instead of writing them on the board would make it clearer for students.

Q6.A was considered to be the best because it incorporates the idea of checking whether students understood the intentions, as it is well-known that students do not always interpret what teachers say (or write) the way teachers thought they would (Wiliam, 2011). The other strategies could have been used to replace the strategy *Mr Fitzgerald* chose, but might not have solved the problem – the teacher would not be able to check how students understood what was expected from them and therefore 'where they were going' in the feedback cycle (Hattie & Timperley, 2007).

Q6.D is considered to be better than Q6.B and Q6.C because the strategy is to give students more details, which could facilitate their interpretation. Q6.B is considered to be better than Q6.C because just communicating verbally can cause even more confusion.

None of these strategies, however, would guarantee that students understood the learning intentions. In Q6.D, writing more could actually have made them even more confusing. In Q6.B, it could be considered as undermining students capacity – it suggests that they are not capable of understanding more than one command. Finally, in Q6.C the word *communicated*[1] was intentionally chosen. In this case, only communicating the intentions verbally to students, as opposed to discussing them *with* students, for example, would have the same result as in only writing them on the board – or even worse considering that they would not be available for students to refer back to when they deemed necessary.

The scenario then returns to the trio *Daniel*, *Carl* and *George* in a conversation with *Mr Fitzgerald* (Figure 7.8).

Facing this problem, Mr Fitzgerald decided to intervene in the conversation.

> **Mr Fitzgerald: How about you, Daniel? What do you think?**
>
> **Daniel: I think we need to divide by half because to divide between two friends is to divide in half...**
>
> **Mr Fitzgerald: So you agree with George... How about you, Carl? Do you agree with your colleagues?**
>
> **Carl: No. I still think that if Claire wants to divide it between two friends, we need to divide it by 2.**

Figure 7.8: *Mr Fitzgerald* intervening in the conversation.

After the dialogue, Q7 is introduced (Figure 7.9).

Question 7. How could Mr Fitzgerald proceed in order to help the group evaluate which way is the most appropriate?

- "I think you have reached an impasse here. You need to talk some more to reach a decision."
- "Let's read the problem again? Can we say that divide in half and divide by half are the same thing?"
- "Why don't you do both calculations and then analyse the answers according to what the problem suggests?"
- "Shall we bring the discussion to the whole group and see if we can come to a conclusion?"

Figure 7.9: Question 7.

In Q7, teachers are required to judge which is the best intervention to help students assess the information they have to be able to solve the problem (an example of KHS). The scores and explanations can be found in Table 7.4.

[1] This was another case that could be included in Section 6.2.3. I made sure to check with teachers what they understood by "*communicating* the learning intentions".

Table 7.4: Question 7 scores and explanations

Option	Score	Explanation
C	4	Teachers identify that *Mr Fitzgerald* is helping students use assessment information to check if their answer is in accordance with which the problem is asking. The help does not give students the answer straight away. The discussion is kept within the group.
B	3	Teachers identify that *Mr Fitzgerald* is helping students use assessment information to check if their answer is in accordance with which the problem is asking. The help, however, gives students the answer straight away. The discussion is kept within the group.
D	2	Teachers do not identify that *Mr Fitzgerald* is not helping students use assessment information to check if their answer is in accordance with which the problem is asking. *Mr Fitzgerald* is still willing to help. The discussion is not kept within the group.
A	1	Teachers do not identify that *Mr Fitzgerald* is not helping students use assessment information to check if their answer is in accordance with which the problem is asking. The discussion is kept within the group.

Q7 comprises two options (Q7.B and Q7.C) in which *Mr Fitzgerald* offers some kind of help to students, and two (Q7.A and Q7.D) in which he does not.

Q7.C was considered to be the best because *Mr Fitzgerald* suggests a strategy in which the work is not done for the students – they still have to calculate and evaluate whether the solution they arrived at is a reasonable response to the problem (Brookhart, 2008; Michalewicz & Fogel, 2013; Polya, 2014)

In Q7.B, although the first part of *Mr Fitzgerald*'s answer could be seen as a good strategy, as it encourages students to analyse what the problem is asking, the second part already tell them what kind of comparison that they should be looking for. On the other hand, students have already read and demonstrated that they understood what the problem is asking. The misconception is not related to how they are interpreting the problem. Furthermore, simply directing students to re-read a text (or part of it) is not an ideal way of helping them with their misconceptions (Kendeou & van den Broek, 2005; Frey & Fisher, 2011).

In Q7.A and Q7.D, the suggestions dismiss students thoughtful process and does not offer proper help. Q7.D is considered to be better than Q7.A because the teacher would at least continue the discussion with the whole group, which would help the trio as well. However, with the information provided, it is not possible to conclude if the whole class was having the same problem or whether this was a problem that only this group was facing. In Q7.A, there is no help at all, as students were already discussing and have not managed to solve the issue among themselves.

Considering the scenario as a whole, teachers are required to analyse whether an activity is in accordance with the learning intentions set out at the beginning. After that, they have the opportunity to interpret one student's mathematical thinking based on a conversation that he is engaging in with a group of students. This is an opportunity that the teacher only had because students were solving the problems in groups. That is, it approaches the importance of talking in the learning of mathematics (MacGregor, 2002; Moschkovich, 1999, 2002; Hodgen & Marshall, 2005; Hodgen & Wiliam, 2006).

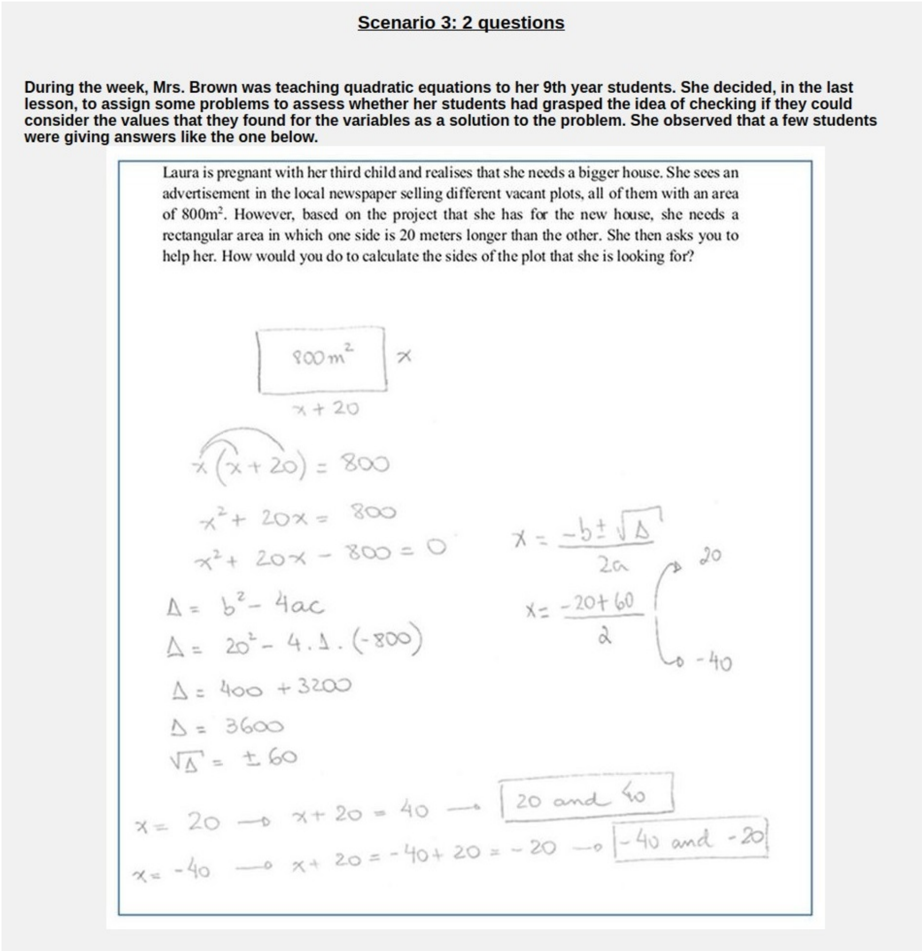
In addition, the scenario also shows that, although *Mr Fitzgerald* has not noticed that his activity was not in accordance with his goals, he would notice it through students' conversation – teachers using assessment formatively to review their own practices.

The scenario also emphasises the importance of articulating clear learning intentions as, in this case, because they were not very clear to students, it hindered rather than helped students to know 'where they were going' (Hattie & Timperley, 2007; Sadler, 2010). It finishes with an example of helping students use the information they have available when students themselves were not able to solve the problem and needed more information (or even a hint).

Finally, this scenario includes a very important aspect of the teaching profession, which FD teachers do not make a good use of: using one another as a resource, or feedback between teachers (Watkins, 2000). This was one characteristic of the MaTFAKi that the researchers considered to be very important. That is, it included some aspects that would also serve as an intervention – to show teachers that there are different possibilities and that they need to talk more and make use of each others' experience to improve their own.

7.3 Scenario 3: Written feedback to some students

Scenario 3 is the shortest scenario, with only two questions. It starts in a similar way to Scenario 1 by presenting *Mrs Brown*'s intentions and an example of how some students were solving the exercise she assigned (see Figure 7.10). The difference from Scenario 1 is that, in this case, it is *the minority* of students that had trouble solving it and there is no question in which the respondent teacher has to interpret students' solution to the problem presented (although the interpretation is somehow required to answer Q8).



Scenario 3: 2 questions

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. She decided, in the last lesson, to assign some problems to assess whether her students had grasped the idea of checking if they could consider the values that they found for the variables as a solution to the problem. She observed that a few students were giving answers like the one below.

Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper selling different vacant plots, all of them with an area of $800m^2$. However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?

Figure 7.10: Beginning of Scenario 3.

Q8 (Figure 7.11) is presented on the same page. Teachers are required to identify the characteristics of effective feedback to judge which is the best feedback to give to students in this situation (an example of KEF).

Question 8. Based on this situation, what would be the best feedback to give to these students?

- Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?
- Congratulations! Your response is almost correct. However, the sides of the plot that you presented are not correct. You need to analyse whether you can consider the values that you found as a solution to the problem.
- Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.
- Congratulations! Your response is almost correct. You just need to check the last part. If the problem asks for the values of the side of a plot, can you accept the negative values?

Figure 7.11: Question 8.

Table 7.5 shows the scores and explanations of each option.

Table 7.5: Question 8 scores and explanations

Option	Score	Explanation
A	4	The feedback starts by noticing what students did, and then points out what they missed. It provides a recipe for improvement at an intermediate level of generality so that students can identify what needs to be done but not so specific that the work is already done for them.
C	3	The feedback starts by noticing what students did, and then points out what they missed. It provides a recipe for improvement that is too vague.
B	2	The feedback does not start by noticing what students did, but points out what they missed. It provides a recipe for improvement that is too vague.
D	1	The feedback does not start by noticing what students did, but points out what they missed. It provides a recipe for improvement that is too specific.

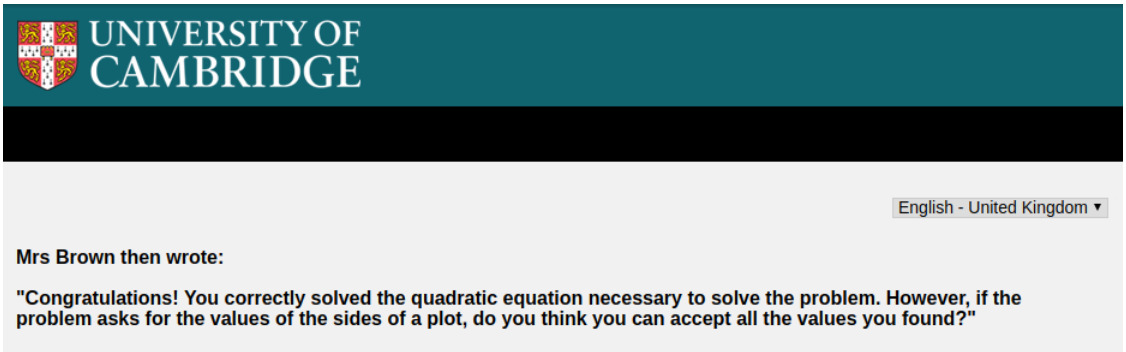
Q8 is comprised of two options (Q8.A and Q8.C) in which the feedback clearly acknowledges what students did; and two options (Q8.B and Q8.D) in which the beginning of the feedback states that the answer was *almost correct* without providing any information about what is it that was 'correct'.

Q8.A is considered the best because the feedback follows by giving students an indication of how they can improve. The feedback, therefore, is specific to the work (Shute, 2008; Stiggins, 2009) and at an intermediate level of generality so that students can identify what needs to be done, but not so specific that the work is already done for them (Brookhart, 2008; Kluger & DeNisi, 1996). Therefore, it presents a recipe for improvement, but also gives students the opportunity to reflect and arrive at their own conclusions about what needs to be done, prompting their thinking.

In Q8.C and Q8.D, although there is some kind of recipe for improvement, it is too vague. Q8.C is considered to be better than Q8.B because the feedback starts by acknowledging what students did correctly, which can reflect positively on their motivation for learning (Brookhart, 1997; Nicol & Macfarlane-Dick, 2006).

Q8.D is considered the least appropriate because, in addition to not showing students what they did correctly, students do not need to identify what needs to be done as the feedback is too specific, and the work is already done for them (Brookhart, 2008; Wiliam, 2011), thereby taking away their learning opportunity. That is, when students finish reading the feedback, they already know that they could not have considered the negative values.

On the next page, Q8.A is reproduced as the feedback that *Mrs Brown* gave (Figure 7.12).


UNIVERSITY OF CAMBRIDGE

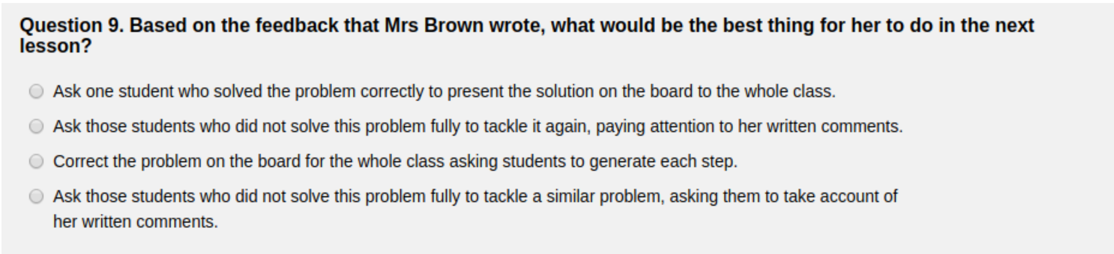
English - United Kingdom ▾

Mrs Brown then wrote:

"Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?"

Figure 7.12: *Mrs Brown* feedback.

In the same page, Q9 is presented (Figure 7.13) and it approaches the importance of giving students the opportunity to act upon the feedback (an example of KCL).


Question 9. Based on the feedback that Mrs Brown wrote, what would be the best thing for her to do in the next lesson?

- Ask one student who solved the problem correctly to present the solution on the board to the whole class.
- Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments.
- Correct the problem on the board for the whole class asking students to generate each step.
- Ask those students who did not solve this problem fully to tackle a similar problem, asking them to take account of her written comments.

Figure 7.13: Question 9.

Table 7.6 shows the scores and explanations of each option.

Table 7.6: Question 9 scores and explanations

Option	Score	Explanation
B	4	Teachers recognise the importance of giving students the opportunity to act upon the feedback. Feedback just given will be taken into consideration. Students will use the feedback on the activity they just solved.
D	3	Teachers recognise the importance of giving students the opportunity to act upon the feedback. Feedback just given will be taken into consideration. Students will use the feedback on a similar activity.
C	2	Teachers do not recognise the importance of giving students the opportunity to act upon the feedback. Feedback just given may not be taken into consideration as the solution will be shown on the board. Students will participate by generating each step for the teacher.
A	1	Teachers do not recognise the importance of giving students the opportunity to act upon the feedback. Feedback just given may not be taken into consideration as the solution will be shown on the board by one student. The other students may not participate.

Q9 is comprised of two options (Q9.B and Q9.D) in which the teacher returns the activity with the feedback for students so they can read it and act upon the the feedback that was written; and two (Q9.A and Q9.C) in which there is some kind of follow up on the board for the whole class – perhaps rendering teachers' written feedback useless.

Q9.B is considered to be the best because it gives students the opportunity to use the feedback they just received in the same activity, whereas in Q9.D this is done with a similar activity, which could be better as a next step to check whether students would be able to transfer the same knowledge to other tasks or situations (Hattie & Timperley, 2007; Kearney et al., 2013).

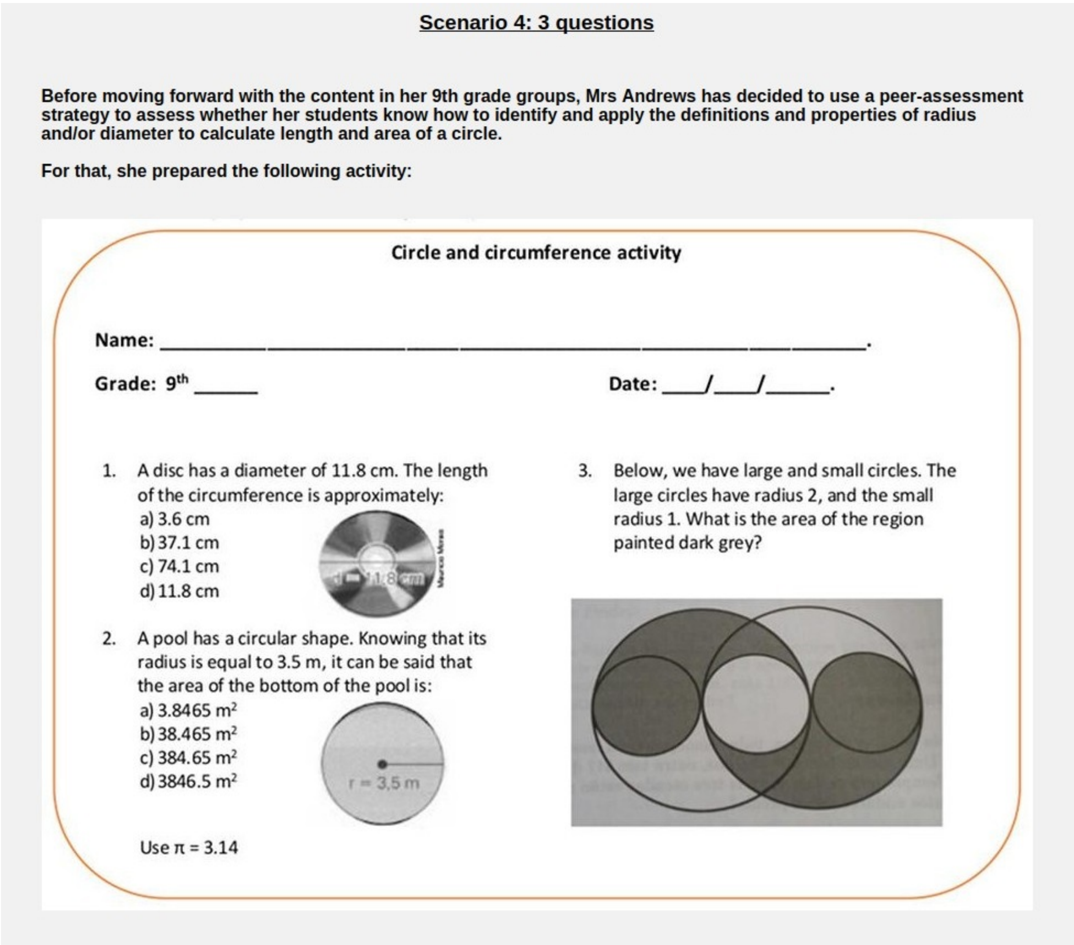
Q9.C is considered better than Q9.A because in Q9.C the teacher will be guiding the solution, while students generate each step, and therefore students' participation will be taken into consideration. In Q9.A, students will only see a correct solution being written on the board, with no kind of explanation or participation. As explained previously, in the Brazilian context, it is quite likely that students will copy the solution from the board and will probably not reflect on their mistakes to improve their learning. In both Q9.A and Q9.C, teachers do consider that discussing misconceptions with the whole class is more suitable when most students are having the same problems (Brookhart, 2008; Hattie & Timperley, 2007; London & Sessa, 2006; Shute, 2008), which is not the case in the situation presented.

In all, the scenario approaches the basis of FA – a teacher using assessment information to provide feedback to students, even if it is the minority of the class that is having trouble (Q8) – followed by the importance of giving students the opportunity to act upon the feedback given (Q9). This process is done through a written activity prepared in accordance with the teacher's learning intention.

7.4 Scenario 4: Peer-assessment activity

In Scenario 4, the idea was to present an example of how teachers could use peer-assessment in their lessons. *Mrs Andrews* prepares a worksheet with some exercises and asks her students to solve, and then provide feedback to a peer. *Mrs Andrews* then uses the evidence to interpret students' learning (Q10). After interpreting, *Mrs Andrews* returns the activity to students, asking them to analyse the feedback given and talk to each other. She realises that students were focussing only on one question. She then asks her colleagues to help her understand why this happened (Q11). At the end, she wants to analyse how the students perceived the learning intentions by asking students to identify them in the activities and feedback (Q12).

The scenario, therefore, starts by stating that *Mrs Andrews* intentionally prepared an activity to be used for peer-assessment. The learning intention is also stated: *to identify and apply the definitions and properties of radius and/or diameter to calculate length and area of a circle*. Teachers are then presented with the exercises that *Mrs Andrews* used (see Figure 7.14).



Scenario 4: 3 questions

Before moving forward with the content in her 9th grade groups, Mrs Andrews has decided to use a peer-assessment strategy to assess whether her students know how to identify and apply the definitions and properties of radius and/or diameter to calculate length and area of a circle.

For that, she prepared the following activity:

Circle and circumference activity

Name: ________________.

Grade: 9th ______ **Date:** ___/___/_____.

1. A disc has a diameter of 11.8 cm. The length of the circumference is approximately:
   a) 3.6 cm
   b) 37.1 cm
   c) 74.1 cm
   d) 11.8 cm

2. A pool has a circular shape. Knowing that its radius is equal to 3.5 m, it can be said that the area of the bottom of the pool is:
   a) 3.8465 m^2
   b) 38.465 m^2
   c) 384.65 m^2
   d) 3846.5 m^2

Use $\pi = 3.14$

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

Figure 7.14: Beginning of Scenario 4.

These exercises are by no means ideal for the activity, but that was also done intentionally so I could ask teachers to analyse them in Q11. In addition, this is the kind of exercises found in textbooks in Brazil, which teachers tend to follow everyday and therefore are frequently found in mathematics lessons.

First, teachers are introduced to *Mrs Andrews*' explanation of what students should do (Figure 7.15).

At the beginning of the lesson, Mrs Andrews handed the activity to students and asked them to solve individually. When they finished, she said:

> **"Now, you will exchange the activity with a colleague. You will analyse their answers, decide whether they are correct, incomplete or wrong and write a comment explaining your rationale. After that, I will collect the activities so I can analyse them myself."**

Figure 7.15: *Mrs Andrews* explanation.

Then, *Chloe*'s and *Derek*'s solutions and comments are presented (Figure 7.16).

Whilst analysing, Mrs Andrews read interesting comments, like Chloe's and Derek's:

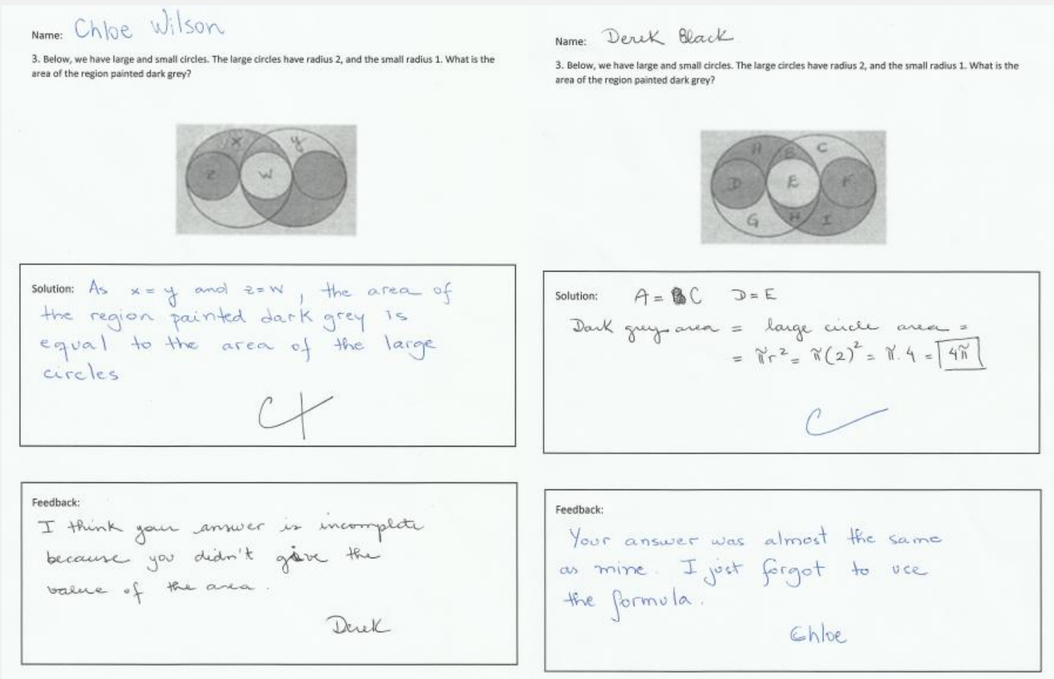


Figure 7.16: *Chloe*'s and *Derek*'s solutions and comments.

The first question in this scenario (Figure 7.17), similarly to Scenario 1, requires teachers to interpret students' learning (another example of KIE). However, in this case, teachers have two different sources to help them: the solution of the exercises and the comment the student wrote to her colleague. The purpose here was to emphasise how valuable giving and receiving feedback can be to the improvement of students' learning, and also to check whether teachers use all assessment information they have available to analyse where their students are in their learning and to make informed decisions in regards to the next steps.

Question 10. Analysing Chloe's solution and her comment on Derek's solution, which is the best way Mrs Andrews can interpret Chloe's learning?

- Even though her first answer was incomplete, Chloe learned when she had the opportunity to assess Derek.
- Chloe did not know how to calculate the area and only commented about the formula because she saw Derek's solution.
- Although Chloe did not present the calculation, her reasoning was correct.
- Chloe knew how to calculate the area, but for some reason did not use the formula.

Figure 7.17: Question 10.

Table 7.7 shows the scores and explanations of each response option in Q10.

Table 7.7: Question 10 scores and explanations

Option	Score	Explanation
A	4	Both *Chloe*'s solution and comment are taken into consideration. With the information provided, it is possible to interpret that *Chloe*'s first answer was incomplete and that she learned when assessing *Derek*.
C	3	Only *Chloe*'s solution is taken into consideration. With the information provided, it is possible to interpret that *Chloe*'s reasoning was correct.
D	2	Only *Chloe*'s solution is taken into consideration. With the information provided, it is not possible to interpret that she knew how to calculate the area.
B	1	Both *Chloe*'s solution and comment are taken into consideration. With the information provided, it is not possible to interpret that she did not know how to calculate the area.

Q10 is comprised of two options (Q10.A and Q10.C) in which teachers' interpretation could be accepted and two (Q10.B and Q10.D) in which the information available is deemed insufficient to arrive to those conclusions.

Q10.A was considered to the be the best because teachers are taking the solution and comment into account. Considering only the solution, the interpretation could be different. However, because *Chloe* wrote the comment, it shows that she realised her mistake when she had the opportunity to assess *Derek*, which might not have happened if the teacher was assessing them. In this case, it shows the importance of peer-assessment in the improvement of learning (Black et al., 2003; Sebba et al., 2008; Yorke, 2003), and an active involvement; allowing the student to practise making judgements not only on her colleague's work, but on her own (Hodgen & Wiliam, 2006). In this case, there is no need for a further intervention from the teacher as the student already identified 'what was missing'.

Q10.C, on the other hand, only takes into account *Chloe*'s answer. Teachers, in this case, are not considering all assessment information available to make an informed decision about the next steps. If that is their interpretation, they would need to intervene to help *Chloe* identify 'what was missing', which would actually be unnecessary as the opportunity of assessing *Derek* already fulfilled the teacher's role in helping *Chloe* to identify her mistake.

In Q10.B and Q10.D, it is not possible to confirm either affirmation, *Chloe did not know how to calculate the area* or *Chloe knew how to calculate the area*, as the information presented is not enough. However, in Q10.A and Q10.C, it is possible to affirm with certainty that *her first answer was incomplete* or that *her reasoning was correct*.

Q10.D is considered to be better than Q10.B because it considers that *Chloe* knew how to calculate the area, as she developed a good part of the solution to the problem, in addition to giving to the option a more positive connotation. Q10.B could be seen as having a negative view of the student taking advantage of the situation.

The scenario goes on with *Mrs Andrews* returning the activities back to students after her analysis (Figure 7.18).

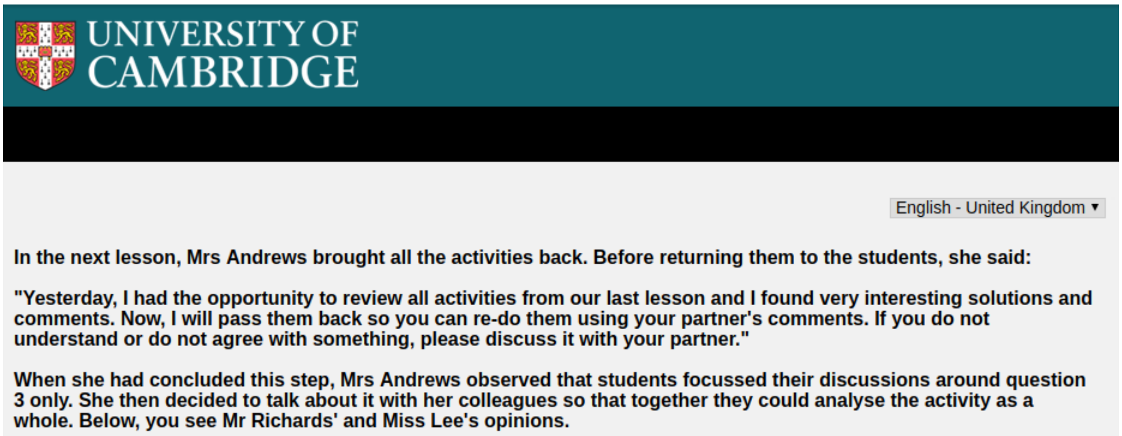


Figure 7.18: *Mrs Andrews* bringing the activities back.

As can be seen, in addition to including an example of feedback coming from a partner, the importance of the teacher inputs is also highlighted. That is, the implementation of peer-assessment also requires teacher participation (Black et al., 2003; Sadler, 2010; Wiliam, 2011). However, as *Mrs Andrews* identified a drawback in the activity, she brings the issue to be discussed with her colleagues, who each give her different explanations (Figure 7.19).

Mr Richards: I think that question 3 generated more discussion because it is more open and allows different solving strategies.

Miss Lee: I, in turn, think it is because of questions 1 and 2 and the fact that they are multiple-choice. The student will select an option and that is it. That is why I do not give my students multiple-choice questions.

Mr Richards: But questions 1 and 2 were in accordance with Mrs Andrews's intention - to assess whether her students know how to use radius and diameter to calculate length and area.

Miss Lee: I do not agree because in question 2, actually, it is only assessing whether the student knows the position of the decimal point when multiplying decimals.

Figure 7.19: *Mr Richards* and *Miss Lee* explanations.

After that, in Q11 (Figure 7.20), teachers are required to evaluate *Mr Richards*' and *Miss Lee*'s opinions in relation to the activity as a whole (another example of KAM). Differently from Q4, where teachers were required to give their opinion, here they are required to judge the opinion of others.

Question 11. Which is the best way of analysing their contributions?

- Mr Richards presented a relevant analysis because he approached both the structure of question 3 and Mrs Andrews's goals.
- Miss Lee presented an important analysis because she highlighted the problem with the format of questions 1 and 2, which explains the attention to question 3.
- Both presented a significant analysis, as they both included important elements in their speeches.
- Both presented an incomplete analysis, which was insufficient to explain the problem presented.

Figure 7.20: Question 11.

The idea with Q11 was to assess whether teachers identify the problem with questions 1 and 2 in which, although they are comprised of some form of real-life application, students are only required to apply the formulae and therefore they represent pseudo-contextualisation.

Table 7.8 shows the scores and explanations of each response option.

Table 7.8: Question 11 scores and explanations

Option	Score	Explanation
A	4	*Mr Richards*' analysis include both the structure of the questions and *Mrs Andrews*' goals. *Mr Richards*' analysis that question 3 is more open and allows different solving strategies is true.
C	3	Both *Mr Richards*' and *Miss Lee*'s analysis are considered to be relevant. Teachers do not consider that *Miss Lee*'s analysis has some problems (discussed in the next option).
B	2	*Miss Lee*'s analysis only considers the structure of the questions. Her analysis that with multiple-choice questions students will select an option and that is it, may not be true in all cases. Her analysis that question 2 is only assessing whether the student knows the position of the decimal point when multiplying decimals only applies if students know that the radius will be squared.
D	1	Both *Mr Richards*' and *Miss Lee*'s analysis are considered to be incomplete. This option contradicts option A which specifies that *Mrs Andrews* explanation can be considered an acceptable explanation.

Q11 comprises one option (Q11.A) in which *Mr Richards*' opinion is analysed and another (Q11.B) in which *Miss Lee*'s is. In the other two options (Q11.C and Q11.D) their opinions are taken together. Q11.C considers that both teachers had good explanations and Q11.D, on the contrary, considers that neither explanation was enough.

Q11.A is considered to be the best because, as it is included in the option itself, it takes into consideration the teacher's goals and the structure of question 3, which is more open and allows different strategies. However, in Q11.B only the structure of questions 1 and 2 are approached.

In Q11.B, *Miss Lee*'s explanation does not explain the real problem as a well-written multiple-choice question can provide teachers with very good insights about their students' understanding, but that will depend on how the response options were developed. Although in the examples included in *Mrs Andrews*' activity, students indeed are required to *know the position of the comma when they multiply decimals* to assess which of the options is correct, they would first need to know how to calculate the area, which, although not ideal, is in accordance with *Mrs Andrews*' goals.

Q11.C is considered to be better than Q11.B because it takes *Mr Richards*' opinion into consideration, and *Miss Lee*'s, which although not completely true, could be considered to have included *important elements*. Q11.D is considered to be the least appropriate because it is too general and dismisses both *Mr Richards*' and *Miss Lee*'s comments.

The scenario continues with a short explanation of *Mrs Andrews*' intentions and

Q12 (see Figure 7.21). In this question, the learning intentions (an example of KLI) are approached differently from the other scenarios – it is at the end of the activity. In this case, the idea is to introduce teachers to another way of addressing the learning intentions, as a way of checking how students perceived what they have been learning, which can prove to be very different from the teachers' intention, and can therefore provide teachers with valuable information about students' learning (Wiliam, 2011).

At the end of the process, Mrs Andrews felt that she should verify whether students had understood what the learning intentions were, using the peer-assessment activity as a reference point.

Question 12. Which would be the best instruction Mrs Andrews could give to the students?

- "Do you remember what we have been learning about measuring circles? I will return the activity from the last lesson so you can have a look at the exercises and feedback and identify what we learned."
- "Do you remember the learning intentions from the last few lessons? I will write them on the board and pass back the activity from the last lesson so you can identify them in the exercises and feedback."
- "Can you tell me what we have been learning about measuring circles? I will ask you to use the activity from the last lesson and the feedback in order to remember. After that, we will write the ideas on the board."
- "Can you tell me what content we have been learning in the last lessons? I will give you the activity from the last lesson so you can review the exercises and the feedback and try to identify what our learning intentions were."

Figure 7.21: Question 12.

Table 7.9 shows the scores and explanations of Q12.

Table 7.9: Question 12 scores and explanations

Option	Score	Explanation
C	4	The beginning of the command is specific, with students being reminded that they have been learning about measuring circles. They will use the activities and feedback *to identify* what they have been learning about measuring circles. There is a follow up opportunity to check if indeed students understood the learning intentions.
A	3	The beginning of the command is specific, with students being reminded that they have been learning about measuring circles. They will use the activities and feedback *to remember* what they have been learning about measuring circles. There is no follow up opportunity to check if indeed students understood the learning intentions.
D	2	The beginning of the command is not specific about measuring circles. Students will use the activities and feedback *to identify* the learning intentions. There is no follow up opportunity to check if indeed students understood them.
B	1	The learning intentions will already be written on the board, so there is nothing for students to identify in the activities and feedback. There is no follow up opportunity to check if indeed students understood them.

In Q12, all options state what *Mrs Andrews* wants students to do after she returns the activities to them. In Q12.A and Q12.C, students are reminded about the main topic 'measuring circles' and in Q12.B and Q12.D the initial question only mentions 'learning intentions', which depending on how used students are to hearing this, can make no sense to them and therefore will not provide any information about what they are supposed to do.

Q12.C is considered to be the best because there is a follow up after students had the opportunity to have a look at the activity. They will analyse by themselves, but will also bring the discussion to the whole group. So, based on their comments, *Mrs Andrews* will have the opportunity to understand how students made sense of the learning intentions (Wiliam, 2011; Wiliam & Leahy, 2015), but will also have the opportunity to come up with an agreed version to make sure everyone is on the same page.

Q12.B is considered the least appropriate because *Mrs Andrews* will write on the board to tell students what they have been learning and the activity will lose its purpose. Once the teacher tells students, there is nothing else to be identified and it will not reflect how students perceive the learning intentions; the 'view' will be from the teacher.

Q12.A is considered better than Q12.D because it very similar to Q12.C; it is more specific and tells students that they were learning about *measuring circles*. Although both do not include a follow up, and therefore, *Mrs Andrews* will be unable to check whether and how students understand what they have been learning, Q12.D is vaguer than Q12.A.

In general, Scenario 4 introduces (or reinforces) to the respondent teachers the importance of peer-assessment activities in helping students to improve their learning and as a source for the teacher to identify what their students have or not learnt. Although using peer-assessment is important, the scenario also highlights that not having well-structured and well-planned activities might mean that the teacher is unable to achieve the intended purpose. At the end, the scenario also approaches how students perceived the learning intentions that might (or should) have been shared with them previously.

In this scenario, therefore, the important idea of using a colleague (be they another student or a teacher) as a feedback resource is included. The teacher encourages feedback between students, but also uses the activity for her own assessment, through her colleagues' comments.

7.5 Scenario 5: Using summative assessments with formative purposes

Scenario 5 is an adaptation from a father's testimony in a blog[2]. It is comprised of four questions. Teachers are required to analyse *Mr Hickman*'s assessment method to explain it to a parent (Q13) and then explain to students how they could make a good use of it (Q14). Next, they are required to choose the best written feedback to give to the student (Q15), who did well in the test, but still presented some aspects in which they could improve; and also to give them some advice of how they could use the feedback that has just been written (Q16). The idea of Scenario 5 was to present FD teachers with an example of how they could use summative assessment for formative purposes.

The scenario starts with a small introduction about *Amy* and a conversation between her and her father (Figure 7.22).

Scenario 5: 4 questions

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 5 on her first quiz of the term and 6 in the second one.

Her father approached her because he was a little concerned as she usually gets more than that.

Father: "Amy, I noticed that your marks in the last maths quizzes were not as high as they usually are. How can I help you, my darling?"

Amy: "Don't worry, dad. Mr Hickman explained that soon we will have an end-of-term test. He will divide the test into sections with each section representing a previous quiz. If I do better in a particular section than I did in the corresponding quiz, my grade for that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure why Mr Hickman would replace the grades. He then decided to go to school to talk to the teacher.

Figure 7.22: Beginning of Scenario 5.

In Q13 (Figure 7.23), teachers are required to judge which would be the best explanation *Mr Hickman* could give to *Amy*'s father with regards to his assessment approach (an example of KAM). Teachers are required to identify that even though *Mr Hickman* had to give marks, the mark does not need to be the final goal. In this case, the teacher could use different quizzes as opportunities for students to receive feedback, and therefore it is much more likely that students will learn from these and their final mark will better reflect what they have learnt.

[2]Available at http://salemafl.ning.com/profiles/blogs/3-perspectives-on-an-afl.

Question 13. What is the best explanation Mr Hickman could give to Amy's father regarding his assessment procedure?

- "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark throughout the term."
- "Although I give a mark for every quiz, they are designed for practice really. If I don't replace the mark, students will not see a reason to study for the test."
- "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this, the grade on the test is the one which will represent what they have indeed learned."
- "The previous quizzes will help students guide their studying and will help them to learn the content by test time. Their mark will better reflect what they really learned."

Figure 7.23: Question 13.

It could be argued that this strategy is teaching to the test, but actually, such procedures have been described and advocated by some authors (Baxter & Glaser, 1998; Laveault, 2013). They allow information from FA to be taken into account in determining a summative assessment and, reciprocally, to use summative information in a formative manner to support learning.

In this question, one of the options includes the idea of *continuous recovery* as an opportunity to recover *the mark*. This was deliberately done because, through my experience in working in different schools and as a teacher trainer, I had the impression that many teachers had this interpretation and therefore it was an opportunity to confirm (or refute) this impression. This is an example in which my experience as a mathematics teacher was beneficial to the development of the MaTFAKi and to the validity of the results.

Table 7.10 shows the scores and explanations of Q13.

Table 7.10: Question 13 scores and explanations

Option	Score	Explanation
D	4	The quizzes are being used formatively as feedback to identify the areas where students are struggling. The mark is referred to as being a reflection of what students have learned through focussing on those areas.
C	3	The quizzes are being used for practice. The mark is referred to as being a better representation of what students have learned as a result of practising the questions.
A	2	The quizzes are being used as 'continuous recovery'. The purpose is to recover the mark throughout the term by giving students various tests.
B	1	The quizzes are being used for practice. The purpose of the mark is to encourage students to study more for the test.

Q13 is comprised of two options (Q13.C and Q13.D) in which learning is the reason of his assessment approach and two options (Q13.A and Q13.B) in which the focus is on getting a better mark. Q13D is considered to be the best because there is specific reference to the quizzes as feedback to guide students studies, whereas in C the quizzes are *just for practice*.

Q13.D was considered to be the best because the teacher explains that his assessment method will be used formatively and the purpose is to encourage students to use assessment information for their own learning (ARG, 2002; Wiliam, 2011). That is, the teacher clearly explains that what matters to him is that *Amy* will have the opportunity to learn based on the feedback from the previous quizzes, and therefore, the quizzes are seen as learning opportunities: points along the journey, rather than the destination. Furthermore, the mark in the test will be a better reflection of learning progress. Students will be able to see how the work along the way helped bring them to that point, showing a connection between feedback and improvement.

Q13.A was considered to the be better than Q13.B because it represents a reality of FD schools, as teachers are required to provide formal recovery opportunities for students. The mistake is to think that these opportunities should be given through tests and that the purpose is only to recover the mark. Although Q13.B and Q13.C start in almost the same way, the second part of Q13.C makes it a better approach then Q13.B because the grade *will represent what they have indeed learned* and in B the mark is the reason to *study for the test*.

The scenario moves straight away to Q14 (Figure 7.24) in which teachers are required to judge how to advise students to use the quizzes (an example of KHS).

Question 14. Which is the best way to advise students to use the tests to improve their learning?

- "You should revise for the test by looking over your old quizzes to identify what you have struggled with, and work more on those areas."
- "You should look for help to revise for the test by going over your old quizzes to learn how to solve the questions that you missed and also re-do those that you already got right at the first attempt."
- "You should re-do all the quizzes many times to practice for the final test of the term. This way, you will for sure have a better mark."
- "You should go over all the past quizzes and re-do all exercises to make sure that you don't miss any."

Figure 7.24: Question 14.

Table 7.11 presents the scores and explanations of Q14.

Table 7.11: Question 14 scores and explanations

Option	Score	Explanation
A	4	Students are advised to go over the old quizzes. The purpose is to identify the areas in which they struggled and focus on those. The focus is on learning the content.
B	3	Students are advised to look for help and go over the old quizzes. The purpose is to learn how to solve those questions they missed, but also to practise the ones they got right already. The focus is on getting the questions right.
D	2	Students are advised to go over all the past quizzes. The purpose is to re-do all exercises to make sure they do not miss any. The focus is on getting the questions right.
C	1	Students are advised to re-do all the past quizzes many times. The purpose is to practise for the test. The focus is on getting a better mark.

Q14 is comprised of four options which approach different elements. Q14.A is considered to be the best because the teacher makes clear that it is the learning that matters at the same time as recognising that students should optimise their review strategies to improve their learning (Black et al., 2003). If students can recognise what they do not know, they can spend their maximum amount of time and effort reviewing that specific content instead of focussing on content that they have already mastered. Therefore, the teacher is not only focussing on the questions themselves, but on the *areas* where students were struggling. It emphasises the content rather than solving the questions correctly. The teacher is providing information to help students to progress (Earl & Katz, 2008).

In Q14.B and Q14.D, the focus is on re-doing all the questions on the quizzes instead of learning the content. Q14.B is considered to be better than Q14.D because students are advised to look for help. This could be from the teacher himself, who, in turn, could help students to focus on their weaknesses.

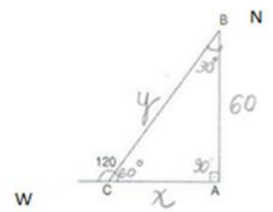
Q14.C is considered the least appropriate because the focus is on re-doing the quizzes *many times* to get a better mark. It gives the impression of mechanical learning, in which by doing the same thing many times is enough to learn the mathematical content and to overcome the difficulties they had at the beginning or assimilate what they have done well.

In Q14.B, Q14.C and Q14.D, as students will not focus only on the concepts that they need to learn, they will not optimise their strategies. The focus is more on knowing how to solve the questions rather than learning the concepts or overcoming their mathematical misconceptions (Hodgen & Wiliam, 2006).

To move on to Q15, there is a small explanation presented to teachers, followed by *Amy*'s solution (Figure 7.25).

After the test, Mr Hickman observed that Amy had overcome the difficulties presented in the first quiz, but could still improve in some aspects of the content of the second quiz. Below, there is her solution to one of the questions about trigonometric identities:

5. A small plane left from town A to a town B, 60 kilometres distant from A towards the north. Due to an orientation problem, the pilot mistakenly followed westward. When he realised it, he corrected the route, turning right 120 degrees at a point C, so that the path, along with the path that should have been followed, formed approximately one triangle rectangle ABC as shown in the picture.



Based on the picture, the distance, in kilometres, that the plane flew from A to B through C was:

a) $30\sqrt{3}$
b) $40\sqrt{3}$
c) $60\sqrt{3}$
d) $80\sqrt{3}$
e) $90\sqrt{3}$

$x+y$?

$Sen\, 60° = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60 \cdot 2$

$y = \frac{120}{\sqrt{3}}$

$tg\, 60° = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$

Figure 7.25: *Amy*'s solution.

In Q15 (Figure 7.26), teachers are required to judge which is the best feedback to give to *Amy* on this exercise (an example of KEF). Although the feedback is being given again after a written activity (as in Scenario 3), this time it is after an end-of-term test.

Given the situation, Mr Hickman decided to write her feedback to help her identify what was missing.

Question 15. Which would be the best feedback to give Amy?

- Well done, Amy! Your reasoning is correct. Now, you need to rationalise the denominator so you can find one of the answers provided.
- Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?
- Well done, Amy! You understood that you needed to use the trigonometric identities to solve the problem. Now, you need to rationalise the denominator so you can find one of the answers provided.
- Well done, Amy! Your solution is almost correct. Can you think of what else you need to do in order to find one of the answers provided?

Figure 7.26: Question 15.

Table 7.12 presents the scores and explanations of Q15.

Table 7.12: Question 15 scores and explanations

Option	Score	Explanation
B	4	The beginning of the feedback makes it clear that *Amy* used the trigonometric identities correctly. The question at the end helps her to identify that there is something else that needs to be done with the final answers, but the work is not done for her.
D	3	The beginning of the feedback does not make it clear that *Amy* used the trigonometric identities correctly. It says that her solution is almost correct. The question at the end helps her to identify that there is something else that needs to be done with the final answers, but the work is not done for her.
C	2	The beginning of the feedback makes it clear that *Amy* used the trigonometric identities correctly. She does not need to identify what is needed next because the teacher already tells her that she need to rationalise the denominator.
A	1	The beginning of the feedback does not make it clear that *Amy* used the trigonometric identities correctly. It refers to her reasoning being correct. She does not need to identify what is needed next because the teacher already tells her that she need to rationalise the denominator.

In Q15, there are two options (Q15.B and Q15.D) in which the teacher provides *Amy* with a recipe for improvement in a way that will make her search for answers to be able to finish the question (prompting her thinking); and two (Q15.A and Q15.C) in which this recipe is too direct and already tells *Amy* what to do – she will do what the teacher suggested without thinking it through.

Q15.B was considered to be the best because the feedback starts by noticing what the student did correctly and therefore there is enough information to answer the question 'How am I going?' in the feedback cycle. In addition, the question at the end prompts student thinking, showing that what she has done so far is correct, but she needs to do something else. Therefore, the recipe for improvement (Wiliam, 2011) is quite clear in relation that there is nothing wrong with her answer, but she still needs to do *something else* – 'Where to next?' However, the work is not done for her as she will have to figure out (Hattie & Timperley, 2007) what is the *something else* that she needs to do. The difference between this option and Q15.D is that in D the teacher only acknowledges that the solution is *almost correct*.

Q15.C is considered to be better than Q15.A because it specifies that *Amy* correctly used the trigonometric identities, and therefore, provides the answer to the question 'How am I going?'; whereas in Q15.A, this answer is not provided. In both options, the recipe for improvement is too specific, with the work being already done for the student (Brookhart, 2008). The question 'Where to next?' was already answered for her and she does not have to engage with her own learning and with what she needs to do to improve.

After that, to introduce the last question of this scenario, the feedback that *Mr Hickman* gave is presented, followed by a small explanation (Figure 7.27).

Mr Hickman wrote the following feedback:

> **Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?**

When returning the test, he called her to his table to talk to her.

Figure 7.27: *Mr Hickman*'s feedback.

In Q16 (Figure 7.28) teachers are required to choose what to say to *Amy* when handing her the test with the feedback (an example of KCL).

Question 16. What is the best thing for Mr Hickman to say when returning the commented test to Amy?

- "You have done well, Amy, because you did much better in the test. You will see that I wrote you some comments so you can learn even more."
- "Amy, you will see that I wrote some comments in some of your questions because you still displayed some difficulties in some aspects. I suggest you to use these comments to try to solve these questions."
- "Well done, Amy! You got a great mark in the test. Now I suggest you read the comments so you can try to solve the questions that you missed."
- "Amy, you progressed a lot compared to the quizzes, but you can still improve in some aspects. I suggest you read my comments to understand what is missing in some questions and try to complete them."

Figure 7.28: Question 16.

Table 7.13 shows the scores and explanations.

Table 7.13: Question 16 scores and explanations

Option	Score	Explanation
D	4	*Mr Hickman* acknowledges that *Amy* improved, but is clear that there are some aspects that she can still improve. The advice focusses on using the feedback to understand what was missing and complete some questions.
B	3	*Mr Hickman* does not acknowledge that *Amy* improved, and highlights that she still displayed some difficulties. The advice focusses on using the feedback to try to solve the questions where she still presented difficulties.
A	2	*Mr Hickman* acknowledges that *Amy* improved because she did much better in the test. The advice is too vague and focusses on using the feedback to learn even more.
C	1	*Mr Hickman* acknowledges that *Amy* got a great mark in the test. The advice focusses on using the feedback to solve the questions she missed.

Q16 is comprised of two options (Q16.B and Q16.D) in which the teacher makes clear to *Amy* that she displayed difficulties or could improve *in some aspects*. That is, that she did well, but missed a few (or small) things; and two in which the suggestion is too general (Q16.A) or focusses on the mark (Q16.C).

Although in Q16.B, the teacher makes clear that the difficulties were only in *some aspects*, Q16.D was considered to be the best because the teacher first acknowledges that the student improved, but that there is still room for improvement. Then, it follows by giving advice on how this improvement can happen. Furthermore, it is clear that the questions were not incorrect, but incomplete. Therefore, it somehow acknowledges that the student is on the right path. In addition, the teacher shows that his comments were made to help her *to understand what is missing* – it identifies the connection between feedback and improvement (Wiliam, 2011). In Q16.D, therefore, the teacher says that *Amy* improved a lot and that she should use the feedback to *complete* the questions where she missed *something*. Whereas in Q16.B, it says that she should try *to solve* the questions as if she had not solved them already.

Q16.B was considered to be better than Q16.A because Q16.A is in the format of praise, and not related to the task. Research has suggested that praise works better with struggling students (Kluger & DeNisi, 1996, 1998), which is not the case with *Amy*, as explained at the beginning of the scenario.

Although Q16.A is too general and does not give any specific direction about what *Amy* should do, it was considered to be better than Q16.C because the praise refers to the improvement she made whereas in Q16.C it is directed to the mark she got. As Dweck (1986, 2006) pointed out, it helps to build a positive attitude in students when teachers show them that they can learn through effort, and has a positive impact on students' motivation (Yin, 2005).

This question shows the respondent teachers that even if other opportunities were given (with all the quizzes and the test), the teacher can still use a "quick-and-quiet feedback" (Brookhart, 2008, p. 49) to help students improve their learning and provide them with an opportunity to close a gap that still has not been closed.

Taking Scenario 5 as a whole, it is assessing whether the respondent teachers can recognise the best way of explaining their assessment method to a parent and how students could make good use of it. After realising that a student is still facing some difficulties, teachers have to choose the best feedback to give and then how to advise this student to make good use of the feedback just received.

As classroom-based summative assessment plays a very important role not only in mathematics lessons in the FD, but in the entire country (Camargo & Ruthven, 2014; Camargo, 2015), this scenario shows an interesting example of how teachers can use these assessments with formative purposes to help their students improve their learning and as feedback to guide their studies.

7.6 Scenario 6: Oral feedback during revision

Scenario 6 was developed to represent a situation of oral feedback. It was adapted from the Knowledge Quartet website[3].

In Scenario 6, *Mrs White* wants to review some content with her students and decides to ask students to develop some questions for this. She has to decide on the best way of devising the criteria for those questions (Q17). Next, after *Mrs White* has started solving one of the question with the whole group, one of her students (*Paul*) makes a comment, which she needs to analyse (Q18) and give feedback on (Q19). The scenario ends with *Mrs White* advising *Paul* and checking whether her feedback helped him move on with his reasoning (Q20).

Therefore, the scenario starts by explaining to the respondent teachers *Mrs White*'s intention of revision and how she is going to approach that: by asking students to develop some problems to be solved together with her during the lesson (Figure 7.29).

[3] Available at http://www.knowledgequartet.org/315/rci-scenario-5/

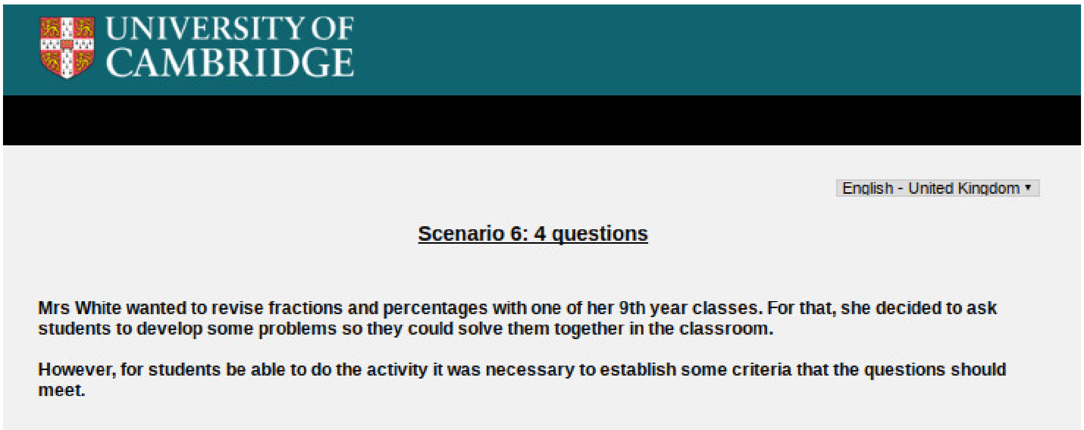


Figure 7.29: Beginning of Scenario 6.

Q17 (Figure 7.30) was developed to assess whether the respondent teachers recognise the importance of including students in the process of devising the success criteria (an example of KLI).

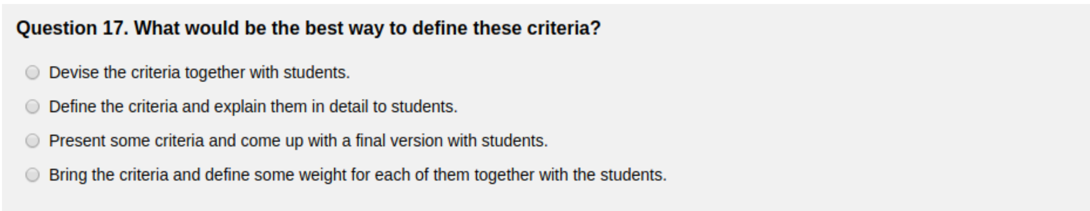


Figure 7.30: Question 17.

Table 7.14 presents its scores and explanations.

Table 7.14: Question 17 scores and explanations

Option	Score	Explanation
C	4	The criteria will be devised together with students, and the teacher provides a starting point.
A	3	The criteria will be devised together with students, and the teacher does not provide a starting point.
D	2	The criteria will be written by the teacher. Students will participate in defining some weight for each of them.
B	1	The criteria will be written by the teacher. Students will not participate.

Q17 involves four different strategies to deal with the success criteria which range from completely participation of the students (Q17.A) to no participation of students at all (Q17.B).

Q17.B, therefore, is the least appropriate because it follows a teacher-centred top-to-bottom approach. The teacher would be the one responsible for devising the criteria and simply deliver them to students, who would only passively receive them and comply to what the teacher has decided. Since one well-known barrier is that of students' lack of understanding of the success criteria (S. Clarke & Fisher, 2009; Stiggins, 2011; Wiliam, 2011), just presenting them to students is not seen as an effective strategy as it will be hard for them to understand what is expected from them and therefore, to meet them.

Q17.D is considered to be better than Q17.B because, although the teacher would bring the criteria, students would at least participate by giving some weight to each of them, somehow giving them some decision power. Therefore, there is some kind of students' participation. The idea of assigning some weight to each criterion, would make students reflect upon them and could be an exercise that would result in a better understanding of each of them. It would also be an opportunity for students to shows the importance they would give to each criterion.

In Q17.A and Q17.C, on the other hand, students are co-designing the criteria together with the teacher. As argued by some (S. Clarke & Fisher, 2009; Wiliam, 2011), when students participate it is more likely that they will be able to apply the criteria. Q17.C is considered to be the best because the teacher would set some starting points, but students would have a great deal of participation in devising a final version of them, whereas in Q17.A, there would be no starting point. Q17.C follows the idea that the setting up of success criteria should not be a democratic process, but a collective negotiation (Wiliam, 2011). The teacher provides students with a starting point so they have an idea of how these criteria should look. That is, the teacher has some control, but students are actively involved in the process.

The criteria are then presented to follow a reasonable flow in the scenario (Figure 7.31). Teachers did not have to do anything on this page and they will not use the criteria to answer the next questions.

Together with the students, they defined the following criteria:

The students should develop, in pairs, 4 questions meeting the following criteria:

1. The questions should approach the content of fractions and/or percentages;
2. The questions should be contextualised;
3. The question can be either open-ended or multiple-choice;
4. You should present at least one solution for each question.

There is no question in this page. Click on the arrows below to continue.

Figure 7.31: Success criteria.

The scenario continues with one of the exercises chosen by *Mrs White* written on the board (Figure 7.32).

In the next lesson, she decided to select some of those questions to solve with the students on the board.

She started with the following:

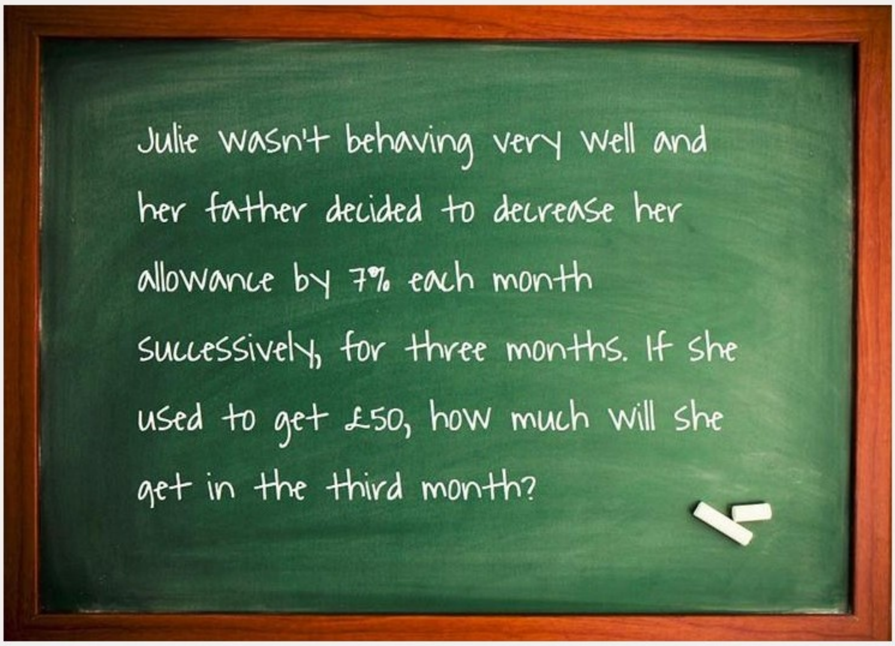


Figure 7.32: Exercise chosen by *Mrs White*.

After, a discussion between the teacher and some students is presented (Figure 7.33).

One student began participating:

Ana: We need to find 7% of 50 and then multiply by 3.

Mrs White: And what do you need to do first?

Ana: Take it away from 50.

Mrs White records the two steps on the board saying:

"This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion.

Paul: Instead of timesing by 0.07 loads of times, we could times by 0.21.

Figure 7.33: *Mrs White* discussing the solution with some students.

In Q18 (Figure 7.34), the respondent teachers are required to interpret *Paul*'s answer (an example of KIE).

Question 18. How could we best interpret Paul's answer?

- Paul is following the teacher suggestion and doing the process again and again.
- Paul is focusing on the 0.07 and the 3, without taking account of the situation.
- Paul is multiplying by 3 instead of using the power of 3.
- Paul has forgotten to take it away from 50.

Figure 7.34: Question 18.

Table 7.15 presents the scores and explanations of Q18.

Table 7.15: Question 18 scores and explanations

Option	Score	Explanation
B	4	This interpretation is possible because when *Paul* mentions *timesing 0.07 loads of time* it shows that he is just manipulating the numbers and is not considering what the problem is asking.
A	3	This interpretation is not possible, because if *Paul* was following the teacher he would calculate decrease after decrease. *Paul*'s answer actually seems to be building up from *Ana*'s answer, who specifically said that they should multiply by 3.
D	2	This interpretation is not possible from *Paul*'s answer. *Paul* will only understand that he needs to take away from 50 once he understands what the problem is asking.
C	1	The interpretation is not possible from *Paul*'s answer. To be able to use the *power of three*, he needs to understand what the problem is asking, and subsequently understand that he needs to take the percentage away from 50 first.

Q18 is comprised of only one option (Q18.B) that includes interpretations that are possible to be made based on *Paul*'s suggestion. The other options include elements that require teachers to make further inferences from what is available to them.

Q18.B, therefore, is considered to be the best because, by suggesting to multiply by 0.21 straight away, it is possible to interpret that *Paul* is not taking the situation into consideration, which requires him to calculate decrease after decrease each month.

Q18.A was considered to be better than Q18.C and Q18.D because Q18.C and Q18.D are explanations that are ahead of *Paul*'s current position, and would have to come together to be considered a plausible explanation. That is, to be able to use the *power of three*, *Paul* needs to understand what *Mrs White* means by *to do the same process again and again*, which also requires *Paul* to understand that he needs to take the percentage away from 50. Therefore, Q18.D was considered to be better than Q18.C because it is a process that *Paul* needs to understand first and it is a necessary step for the solution; whereas *using the power of three* is optional. That is, it is possible to solve the problem without using the power of three, but it is not possible to solve it without taking 7% from the initial amount of 50.

On the next page, with no follow up text, Q19 is introduced (Figure 7.35) asking the respondent teachers to assess which is the best feedback to give to *Paul* based on what he just suggested (another example of KEF).

Question 19. In this case, what would be the best feedback to give to Paul?

- "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"
- "0.21 is 3 times 0.07, right? Can you figure out whether this solution may or may not work in this problem, Paul?"
- "Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case."
- "0.21 may not work, Paul, but I will record your suggestion on the board so we can discuss it later."

Figure 7.35: Question 19.

In this case, the teacher is required to assess the characteristics of an effective oral feedback to make *Paul* realise that what he is thinking will not work, but at the same time, will avoid giving him the answer or discouraging him from engaging in finding another solution.

Table 7.16 shows Q19 scores and explanations.

Table 7.16: Question 19 scores and explanations

Option	Score	Explanation
C	4	The teacher gives *Paul* the opportunity to think about his answer. The teacher asks *Paul* to explain where the 0.21 came from.
B	3	The teacher gives *Paul* the opportunity to think about his answer. The teacher tries to guess where the 0.21 came from.
A	2	The teacher does not give *Paul* the opportunity to think about his answer as she tells him straight away the problem with his suggestion. The teacher gives him the opportunity to think about another solution.
D	1	The teacher does not give *Paul* the opportunity to think about his answer as she already tells him that it *may not work*. The teacher does not give him the opportunity to think about another solution.

Q19 comprises two options (Q19.B and Q19.C) in which the teacher gives *Paul* the opportunity to think about his answer and two (Q19.A and Q19.D) that puts *Paul* off, either by giving him the answer straight away or moving on with the lesson.

Q19.C is considered to be the best because the teacher realises that there is some kind of misconception in *Paul*'s answer and, by asking him to show her how he came up with 0.21, she is making him think and probably he will realise that his strategy might not work. Therefore, the feedback is giving *Paul* the opportunity to think about his answer and improve it. It is specific to *Paul*'s answer (Shute, 2008; Stiggins, 2009) and provides information that can help him identify what needs to be done (Brookhart, 2008; Kluger & DeNisi, 1996; Wiliam, 2011). In this case, therefore, the teacher asks for the student's explanation and then ask him to reason. It will help *Paul* to understand his own thinking.

In Q19.B, however, by saying *0.21 is 3 times 0.07, right?* she is anticipating his thoughts, and not making him explain what he was thinking. In this case, the teacher is not considering the importance of letting the student talk about his strategies, which could, in turn, help him understand his own thought processes (MacGregor, 2002; Moschkovich, 1999, 2002; Hodgen & Marshall, 2005; Hodgen & Wiliam, 2006).

Q19.A is considered to be better than Q19.D because, although the teacher already tells him the answer, she at least gives *Paul* the chance to try another solution, whereas in Q19.D the teacher already dismisses *Paul* and ends the conversation straight away. By doing this, she will not elicit what he was thinking, and is not helping him to understand why his answer is not correct (Hodgen & Wiliam, 2006). In this case, she is also not taking into account the motivational and sense of self-efficacy aspects of assessment (Yin, 2005) as she seems to be dismissing *Paul*, which might show him that it was because he made a mistake.

To move to the last question, a dialogue between *Mrs White* and *Paul* is shown (Figure 7.36).

And the dialogue continued:

Mrs White: Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case.

Paul: The problem says three months. 3 times 0.07 is 0.21.

Mrs White: Yes, I agree with you: 0.21 is 3 times 0.07, but do you think this is what the problem is asking?

Paul: Well... the problem says 3 months, that's why I thought I had to multiply by 3.

Mrs White: It says 3 months, Paul, but it's 7% in each month, and not all in the end of the third month, ok?

Paul: Ah ... I get it. So we need first to take away 7% and then 7% again and 7% again.

Figure 7.36: Dialogue between *Mrs White* and *Paul*.

The MaTFAKi ends with Q20 (Figure 7.37) in which the respondent teachers are required to choose the best way for *Mrs White* to check whether *Paul* has learnt with the feedback given (an example of KCL).

Question 20. In this case, how should the teacher continue the discussion in order to verify if Paul learned with the feedback given?

- "So let's do this: you solve it in your notebook and when I show it on the board, you check if you got the correct answer."
- "Has everybody understood what Paul just said? Let's do it together on the board."
- "Yes. You take 7% of 50. Then, from what is left, you take 7% again and so on. Do you want to solve it on the board so everyone can see it?"
- "Let's do this: you calculate 21% and then you do like that, taking away 7, 7 and 7. When you finish, we can talk about the results."

Choose the following options if you want to see any part of scenario 6 again.

- View the dialogue between the teacher and students.
- View the problem written on the board.

Figure 7.37: Question 20.

Table 7.17 presents the scores and explanations of Q20.

Table 7.17: Question 20 scores and explanations

Option	Score	Explanation
C	4	*Mrs White* takes *Paul*'s suggestion into consideration, but also reinforces the importance of taking it away from 50 first. By inviting him to solve on the board she will be able to check immediately if he really understood.
D	3	*Mrs White* takes *Paul*'s suggestion into consideration. Her suggestion of doing both calculations seems unnecessary as his comment shows that he already understood that 0.21 will not work. By suggesting that they will talk about the results when he finishes, she will be able to check if he really understood.
B	2	*Mrs White* takes *Paul*'s suggestion into consideration. By solving the exercise on the board herself she will not be able to check if he really understood.
A	1	*Mrs White* does not take *Paul*'s suggestion into consideration. By solving the exercise on the board herself she will not be able to check if he really understood.

Q20 is comprised of two options (Q20.C and Q20.D) in which *Mrs White* will be able to check whether *Paul* has learnt from the feedback given, and two (Q20.A and Q20.B) in which she will be unable to check if he learned with the feedback and dismisses him either by asking him to solve by himself and check later or by moving on to showing the correct solution on the board.

Q20.C is considered to be the best because the teacher reminds *Paul* of the importance of *taking it away from 50 first*, but also considers his suggestion. By inviting him to show to the whole group, she is valuing his suggestion, she will be able to check what *Paul* is doing, and at the same time interact with the whole group through his solution.

In Q20.D, although it is a good approach to ask him to try to solve using both strategies, *Mrs White* is somehow ignoring what *Paul* just said. Based on his comment, he already understood that using 21% will not work. It would be a good strategy beforehand, so he could have arrived at the conclusion by himself. However, as she already explained that it is *7% in each month, and not all in the end of the month*, it does not make sense to have him do it again. Q20.D is considered to be better than Q20.B because the teacher will be able to check whether *Paul* learnt from the feedback as they *will talk about the results later*.

Both Q20.B and Q20.A take the discussion away from *Paul*. Q20.B is considered to be better than Q20.A because they will do it *together* and in Q20.A *Paul* will solely check his final answer, and the teacher will actually have no means to verify it. She will provide the correct solution on the board and he will only confirm (or not) whether what he was thinking was the same as presented on the board. The focus is on getting the correct answer.

Considering the scenario as a whole, the respondent teachers have to decide what is the best way of devising the criteria of some questions that *Mrs White* asked students to create. Next, teachers are required to analyse the answer of one of *Mrs White*'s students and the best feedback to give based on the answer just given. The scenario ends with teachers judging what constitutes the best advice Mrs White should give to *Paul* to check whether her feedback helped him to move on with his reasoning.

7.7 Summary

In this chapter, I presented the final version of the MaTFAKi. It was important to explain its rationale and the theory behind each question as this will be the basis for presenting and discussing the findings in the next chapters.

As the MaTFAKi is an index measuring different aspects related to FA, it was essential to include different aspects of each DoK being assessed, as well as different aspects of the teaching and learning of mathematics.

There are many other different ways in which the aspects included and other aspects could have been approached, but it was not the intention of this chapter to discuss these aspects. However, I felt that at this point, the first research question needed a small adjustment so that I could indeed answer it:

> What do secondary mathematics teachers in the Brazilian FD know about FA in general and the idea of feedback in particular, *as measured by the MaTFAKi*?

The reason was that the information being collect about teachers' FA knowledge was specific to the examples included in the MaTFAKi, and therefore I could not make inferences about teachers' knowledge without specifying that it was information 'as measured by the MaTFAKi'.

In the next chapter, I present how I distributed the MaTFAKi and how the data was analysed to be able to answer this and the other research questions.

CHAPTER 8

DATA COLLECTION AND ANALYSIS

The previous chapters explained all the steps followed to develop and validate the MaTFAKi, and provided a detailed account of the final version sent to FD teachers. This chapter explains how the MaTFAKi was distributed to participants and how the data was analysed.

Section 8.1 presents how the MaTFAKi was sent to FD teachers, and compares the sample strategy with the strategy used in the field test (hereafter referred as *BR* teachers). Section 8.2 explains how the data was prepared and analysed, with a brief explanation of the statistical techniques used.

8.1 Sending the MaTFAKi

In principle, I set out to collect data from the entire population of FD secondary mathematics teachers, although I realised that, in practice, I would not receive responses from all of them. As I had access to the official email addresses of all schools, but not to those of teachers individually, I decided to contact all 352 secondary schools in the FD asking principals to forward the email explaining the study and requesting cooperation to all mathematics teachers in the school. To incentivise participation, I offered a prize draw of four vouchers of R$100.00[1] each for an online store of their choice.

Compared with the strategy followed in the field test, the approach was the same – asking the principal to forward the email to teachers – but at that stage, because I was

[1] The equivalent of approximately £25.00.

dealing with a much larger population, a random sample of schools was chosen (see Section 6.3.1) and teachers were not offered any incentive.

To keep teachers' anonymity, I did not include any question that could identify their schools. Therefore, it was not possible to keep track of how many teachers from each school answered the MaTFAKi. However, it was possible to track if an email bounced, and for all 23 emails that bounced, I called the schools and asked for the correct address, so that I could make sure that all schools received the link to the MaTFAKi and all teachers would have an equal chance to receive it. All emails were sent using the Qualtrics® mail system.

The MaTFAKi was available online for two months and various reminders were sent within this period. The characteristics of the respondents are presented in the following chapter. However, it is important to explain at this stage that, although I did not request this, some teachers forwarded the email to other colleagues. I am aware of this because some teachers emailed me asking if they could answer the MaTFAKi. As they were all FD mathematics teachers, and willing to forward the request to other colleagues, thereby helping with the distribution of the MaTFAKi, my answer was positive to all of them. This explains why some teachers from independent schools[2] answered the MaTFAKi and shows that some principals did not forward the email as I both requested and wished. In any case, a 'thank you' email was sent to all schools after the two-month period. Equally, all emails with questions or requests within this period were answered in a friendly way.

8.2 Preparing and analysing the data

Although it is known that the analysis of quantitative data is relatively less time consuming (Cohen et al., 2011), a systematic approach is still needed. This section explains the steps followed in this study.

8.2.1 Data extraction and preparation

As the MaTFAKi was developed using Qualtrics®, it was possible to export all the data directly as a file in '.csv' format to be used in *R*, where all the statistical analyses were performed.

[2]These teachers were later excluded from the analysis as they were outside the population I set out to survey.

The preparation of the data from the closed-response items was done in three parts. The first involved translating all data labels as the MaTFAKi was prepared and answered in Portuguese. The translation was necessary because the labels of charts, or the data from tables would be automatically generated by *R* from the file exported from Qualtrics®. The second was the coding. The majority was automatically done by Qualtrics® before exporting took place. However, some further coding and rescaling of this data was done in *R*, and some filtering was also carried out. Where such action was taken, it will be made clear in the text. The third step was classifying the variables in *R*. In this case, there were nominal (some of teachers' personal characteristics), ordinal (the questions themselves) and interval (total scores) variables.

8.2.2 Data analysis

This section explains how the data analysis was carried out, based on the different types of data, and how the results of the statistical tests will be presented in the text.

Participants' background information included numerical and non-numerical data. Results for the numerical data are summarised using descriptive statistics, with tables and charts used to present them in the text. The non-numerical data comprised information that would better explain teachers' level of education (such as their first degree, the subject of their masters, etc.) and which degree they had if they stated that they did not hold a teaching qualification in mathematics. Grouping and counting of this data was done by hand, with the help of MS Excel.

The non-numerical data required some cleaning, but no coding as such. For both variables, I first checked all the answers and corrected any misspelling. Second, I made sure that the same degree was receiving the same denomination. For example: 'matemática - bacharel' and 'bacharelado em matemática' refer to the same degree: mathematics (without the teaching qualification). Therefore, in this case, I would change both to 'Bacharelado em matemática', which is the more common denomination, and Excel would be able to recognise and count them as the same. As I was not dealing with a large amount of data, this was done by myself, row by row. Tables were generated for each level of education. The results are summarised in Section 9.1.2.

The overall reliability of the 20-item scale with the results of FD teachers was also calculated through ordinal alpha (Zumbo et al., 2007). Although (see Section 6.3) some believe this to be unsuitable for SJTs (e.g. Cabrera & Nguyen, 2001; Christian et al., 2010), reliability analysis based on alpha was carried out because it provides interesting information, especially in terms of the reliability of the scale with each item removed (Table 8.1) and of how the items correlated when considering the scale as whole (Table 8.2).

Table 8.1: Reliability analysis of the 20-item scale when responded to by FD teachers.

raw_alpha	std.alpha	G6(smc)	average_r	S/N
0.55	0.55	0.67	0.058	1.2

Reliability if an item is dropped:					
	raw_alpha	std.alpha	G6(smc)	average_r	S/N
Q1_KIE	.54	.54	.66	.057	1.2
Q2_KEF	.51	.51	.62	.052	1.0
Q3_KCL	.54	.54	.65	.057	1.2
Q4_KAM	.54	.54	.65	.059	1.2
Q5_KIE	.55	.55	.66	.061	1.2
Q6_KLI	.53	.53	.66	.057	1.1
Q7_KHS	.53	.53	.65	.056	1.1
Q8_KEF	.51	.51	.63	.052	1.1
Q9_KCL	.54	.54	.66	.059	1.2
Q10_KIE	.54	.54	.64	.058	1.2
Q11_KAM	.60	.60	.70	.075	1.5
Q12_KLI	.53	.53	.65	.056	1.1
Q13_KAM	.55	.55	.66	.060	1.2
Q14_KHS	.52	.52	.64	.054	1.1
Q15_KEF	.50	.50	.62	.051	1.0
Q16_KCL	.55	.55	.67	.061	1.2
Q17_KLI	.55	.55	.66	.061	1.2
Q18_KIE	.54	.54	.64	.058	1.2
Q19_KEF	.55	.55	.66	.060	1.2
Q20_KCL	.55	.55	.67	.062	1.2

Table 8.1 shows the results of the reliability analysis, **raw_alpha** usually represents alpha based upon the covariances. In this study, it represents 'ordinal alpha' (Gadermann et al., 2012), because it is based on the polychoric correlation matrix; **std.alpha** represents the standardised alpha based upon the correlations; **G6(smc)**, Guttmans Lambda 6 reliability, and **average_r**, the average inter-item correlation.

As can be observed, when the 20-item scale was answered by FD teachers, it resulted in a much lower value ($\alpha = .55$) than the 24-item version ($\alpha = .68$) used in the field test, or the 20-item scale after the retest results ($\alpha = .66$).

Table 8.1 also shows that the reliability of the whole measure would increase considerably if Q11 was removed – from .55 to .60. – and decrease – from .55 to .50 if Q15 was removed. These results can be partially explained by the inverse correlation of Q11 ($r = -.10$, Table 8.2), and the high correlation of Q15 ($r = -.52$), but also by how differently some questions behaved when comparing the FD and *BR* teachers, as will be explained in the next chapter.

Table 8.2: Item statistics of the 20-item scale when responded to by FD teachers.

	r	r.cor	r.drop
Q1_KIE	.35	.29	.206
Q2_KEF	.49	.51	.356
Q3_KCL	.35	.29	.202
Q4_KAM	.31	.25	.159
Q5_KIE	.26	.19	.107
Q6_KLI	.37	.29	.225
Q7_KHS	.38	.33	.239
Q8_KEF	.48	.48	.346
Q9_KCL	.30	.22	.152
Q10_KIE	.34	.30	.198
Q11_KAM	-.10	-.27	-.250
Q12_KLI	.37	.31	.231
Q13_KAM	.29	.22	.143
Q14_KHS	.43	.39	.289
Q15_KEF	.52	.53	.392
Q16_KCL	.26	.17	.111
Q17_KLI	.26	.18	.106
Q18_KIE	.32	.29	.175
Q19_KEF	.28	.20	.130
Q20_KCL	.24	.14	.087

In Table 8.2, **r** represents the correlation of this item with the whole scale; **r.cor**, the item-whole correlation corrected for item overlap and scale reliability; and **r.drop**, the item-whole correlation *for this item* against the scale without this item.

To explore variable structures in the dataset (Field et al., 2012) or to identify underlying dimensions in the data, an EFA[3] was also conducted.

[3]The explanation can be found in Section 6.3.2.

As happened in the field test, the factor analysis did not yield interpretable factors (Chan & Schmitt, 1997; McDaniel & Whetzel, 2005). As the results of the factor analysis from the larger field test sample have already been presented in detail in Section 6.3.2, they are not replicated here. However, it is important to explain that although the result of the Bartlett's test was highly significant ($\chi^2_{190} = 584.48, p < .001$) and the determinant of the correlation matrix (0.02620844) was above the necessary value of 0.00001, the sample size of the main survey was unsuitable for EFA ($KMO =$ .4) (Field et al., 2012; Kaiser, 1970). In this case, it would be necessary to collect more data, but due to the time frame of a PhD project, this was not possible. In addition, when I closed the responses for the MaTFAKi, no more answers were being received. Although reminders were sent, it did not improve the response rate.

Consequently, items will not be grouped by DoK based on the results of the factor analysis, but based on the patterns of teachers' responses (see Section 9.3). However, as each item was developed to assess a specific aspect of a specific DoK, and all the trial phases confirmed that they were indeed assessing the relevant aspect, the DoK will still be referred to.

In terms of the 20-item scale, to have an overall feeling for the data, the mean and standard deviation of teachers' scores were generated for the MaTFAKi as a whole and then compared to the field test sample.

To analyse whether this difference was statistically significant, a Welch two sample t-test was carried out (Fagerland & Sandvik, 2009; Ruxton, 2006). The results are presented in Section 9.2.

The Welch two sample t-test (Welch, 1947) tests the hypothesis that two populations have equal means without assuming that those two populations have the same variance. In addition, it reports a confidence interval for the difference between the two means that is usable even if the variances differ.

It is important to emphasise here that I am aware that calculating the mean scores of ordinal data does not provide fully comparable scales because the intervals between categories are subjective or non-existent (Agresti, 2002). However, I believe that comparing and reporting the mean of total scores in this manner provided a clearer view of where FD teachers stand in relation to their overall knowledge of FA, when using *BR* teachers' performance as a benchmark. In the next chapter, I will provide more information from all questions and also a deeper analysis of the data to support the argument that the FD sample, in general, had a lower performance than the *BR* sample.

For this reason, to compare the difference in total scores not only based on their mean scores, another test was performed. The Mann Whitney U or Mann-Whitney-Wilcoxon (Hettmansperger & McKean, 2010; M. Hollander, Wolfe, & Chicken, 2013) is the test most commonly used to compare rankings, or ordinal values, in two separate sample groups to determine if they reflect the presence of a real difference in the populations they represent. The null hypothesis simply states that there is no systematic or consistent difference between the two populations being compared. The results of the Mann-Whitney-Wilcoxon test is also presented in Section 9.2.

Yet, I was also aware that this kind of total score analysis would not be helpful to understand which aspects of FA teachers already performed well and in which aspects they should improve. Therefore, tables and charts with frequencies and percentages were also generated to provide a clear view of how teachers' responses were distributed for each item. As there are no other studies of this kind in Brazil, the responses from the FD were compared to the responses from *BR* teachers, which were used as a benchmark for the analysis. This comparison was possible because no questions were changed after the field test (although some were dropped on the basis of the test-retest reliability analysis).

To compare how the responses differed, a test of equal or given proportions ('prop.test') was applied (Newcombe, 1998a, 1998b). The 'prop.test' can be used for testing the null that the proportions in several groups are the same, or that they equal certain given values. In this study, it was used to compare the proportion of teachers choosing each score option between the FD and *BR* samples. A low p-value indicates that the proportions probably differ from each other. To facilitate the comparison, I will use the notation $(FD_n = q\%)$, where q is the observed percentage of the FD teachers who chose score n option. *BR* will be used $(BR_n = q\%)$, instead of FD, to report the percentages of teachers from the other states of *BR*azil.

Finally, **to better understand how teachers' knowledge was organised**, and because the factor analysis did not generate interpretable factors, cross-tabulation was employed to analyse the degree of association and homogeneity among all questions that were related and with teachers' demographic data.

With nominal variables, Chi-square tests were applied and p-values were obtained through Monte-Carlo estimation to evaluate the statistical significance of the relationship between the questions, taking into account the degrees of freedom (df).

The Chi-square test (χ^2) is employed to test the difference between an actual sample and another hypothetical or previously established distribution, such as that which may be expected due to chance or probability. However, it can also be used to test the differences and relationships between two or more variables (Field et al., 2012; Key, 1997; Agresti, 2002). The results of the Chi-square test must be compared to the critical values from a Chi-square table, which is given in relation to the number of *df* and the p-value.

For two-dimensional contingency tables, Pearson's Chi-square test is most frequently used, although it is common to use Fisher's exact test for 2×2 tables when the sample size is small (Guo & Thompson, 1989). In my data, it was observed that many of the tables were sparse, i.e. the frequency of some cells did not exceed 5, and the contingency tables, in all cases, were bigger than 2×2. Thus, the Monte Carlo exact test was used, since it takes into account the sparsity and size of the contingency tables, thereby providing more accurate results.

In relation to the p-value, essentially what is being analysed is whether the probability of the result is simply due to chance – to some random factor in the data – as opposed to some systematic difference or trend in the data that would indicate a real finding.

The results of the Chi-square tests are shown in the text as: $(\chi^2_{df} = q, p < \alpha)$, where q is the observed value of the χ^2 statistic with df degrees of freedom and p is the p-value, which is compared to the significance level α. In the social sciences, the minimum level of probability at which a result can be regarded as statistically significant is $p < .05$.

The same was done with regards to the ordinal variables. However, the analysis of the significance was made through a Goodman Kruskal's gamma test, which measures the relationship between two ordinal or ranked variables. In addition to providing the significance, the sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. The possible values range from -1 to +1, but the extreme values can be obtained only from square tables (Sirkin, 2006). Where a Goodman Kruskal's gamma test was used, the results are shown in the text as: $(\gamma = q, p < \alpha)$, where q is the observed value of the *gamma* statistic and p is the p-value, which is compared to the significance level α.

8.2.3 Writing up the report

Once the analysis had been completed, I started writing the report. As explained, the responses from the closed-ended and open-ended questions have been used to characterise the respondents, and to arrive at the findings about their overall knowledge of FA and the specific knowledge being assessed with each question. It was important to run all the analysis before writing the report to have a clear view of which would be the best way to present them.

8.3 Summary

In this chapter, I started by presenting the sample strategy for the main survey and compared this with the field test which was essential for comparing their results. I then moved on to presenting how the data was prepared and analysed. First I presented the results of the reliability analysis and, once that was established and could be considered an acceptable result, I presented the statistical tests that were carried out to assess the FD teachers' total scores and to compare their results with those of *BR* teachers.

Figure 8.1 provides an overview of the process followed to prepare and analyse the data as reported in this chapter.

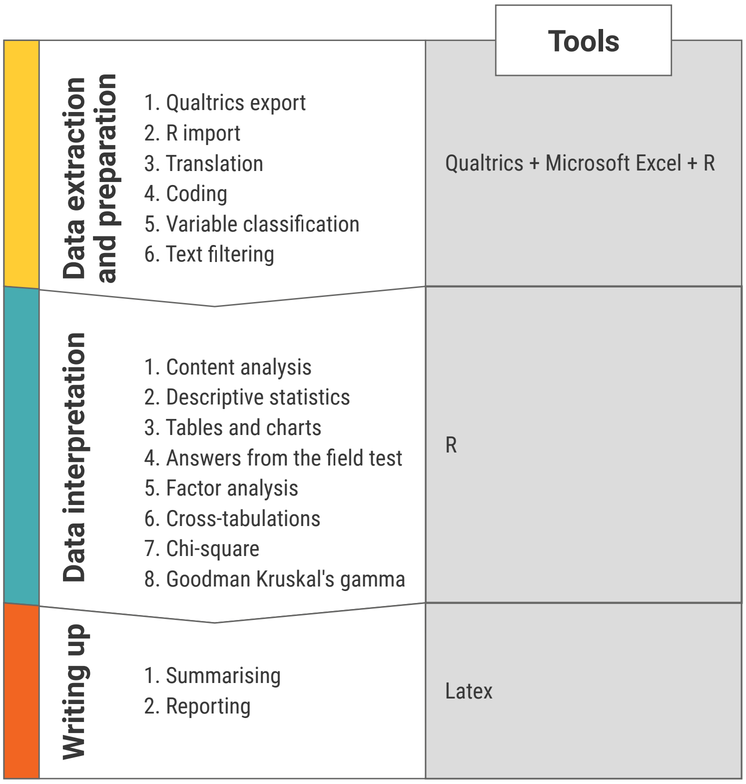


Figure 8.1: Data analysis process.

In the next chapter, I present the main findings of this study.

CHAPTER 9

FINDINGS

The previous chapter presented how the MaTFAKi was distributed to participants and how the data was analysed. This chapter presents the findings of this inquiry.

Section 9.1 summarises participants' personal, academic and employment characteristics. Section 9.2 examines the overall knowledge that FD mathematics teachers have of FA, based on their total score in the MaTFAKi. Section 9.3 presents the results of each questionnaire item and explores the behaviour of some items separately. Section 9.4 compares the results between the sample from the FD and that from the other states in Brazil (*BR*). Section 9.5 explores how some items were related to each other in an attempt to uncover connections among variables.

9.1 Characteristics of the participants

In this section, the purpose is to present the characteristics of the 169[1] FD teachers who participated in this study. Compared to the total sample of the field test this is a smaller sample, but considering that the FD is the equivalent of one state, proportionally, this number is well above the number of responses from any other state that participated in the field test. For example, Sao Paulo, which is the most populated state in Brazil, had 57 participants.

Usually, field tests are done with a small sample and then the main survey with a

[1]This number already excludes teachers that reported teaching only in independent schools and therefore were outside the population I set out to survey (see Section 8.1).

larger sample. In this study, the opposite was done, primarily because the main survey would be conducted in a small population. Therefore, it is not surprising that the number of participants in the main survey is less than the total number of those participating in the field test.

Although smaller, the sample can be considered a representative one, as explained in the following sections. To better characterise the participant teachers, the demographic information was organised into three groups:

1. **Personal characteristics:** age and gender;
2. **Academic characteristics:** level of formal education and teaching qualification;
3. **Employment characteristics:** years of experience, whether they teach in private or state schools, and in which grades they were teaching when they answered the MaTFAKi.

These variables have either nominal or ordinal measurement levels with various categories and scales. To ease interpretation of results, some of these scales were further simplified. For example, 'Age' was reduced from a 6-item scale to a 3-item scale: Younger (39 or less); Middle age (40–49) and Older (50 or more). 'Grades teaching this year' moved from seven options to four: Not teaching this year, Fundamental Education II, Middle Education, or Both[2].

9.1.1 Personal characteristics

In regards to the participants' gender, 56%[3] are male and 43% female. 1% of the participants opted for not stating their gender (see Figure 9.1).

I could not find any study comparing gender distribution amongst mathematics teachers in Brazil or the FD. In the field test sample, however, 51% of the respondents were women and 49% men.

[2]See Table 2.2 for reference.

[3]Due to the size of the sample, percentages will be quoted as whole numbers to avoid spurious accuracy.

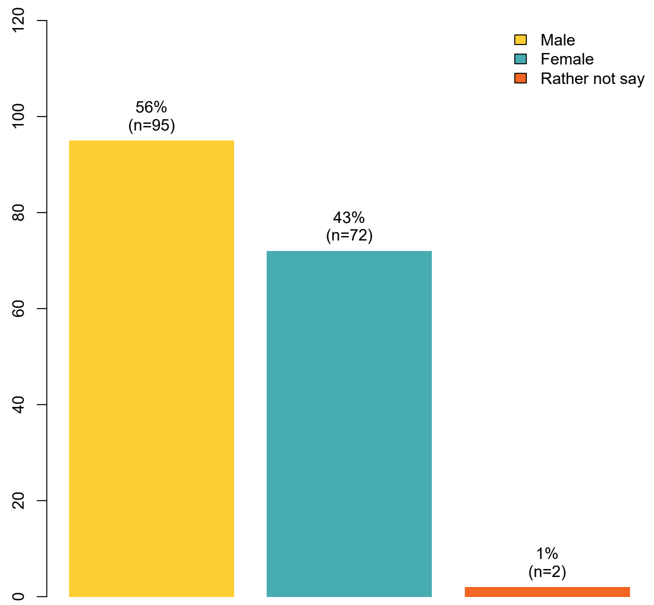


Figure 9.1: Gender distribution.

In regards to their ‘Age’, Figure 9.2 shows that there is a considerable number of participants in every age group.

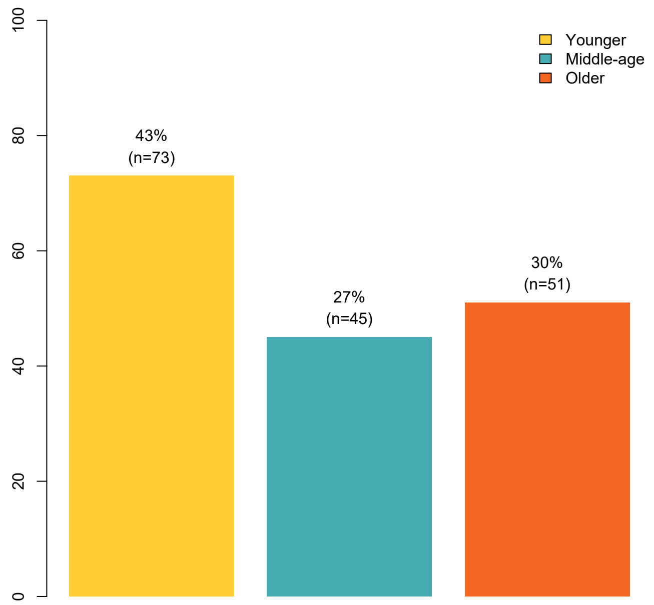


Figure 9.2: Age distribution.

This result is important because, although the age categories used in this study are not linearly defined, every age group has adequate representation in the data. It can be observed that 'Younger' teachers are represented in higher proportions (43%) than the other two age groups (27% and 30% respectively). However, it is difficult to clarify without additional/official demographic information how much this skewness should be attributed to demographical characteristics of mathematics teachers, to the unevenness of the scale, or to a higher response likelihood of 'Younger' teachers. If compared with the characteristics of the pilot sample, 'Younger' teachers were also represented in higher proportion there (see Section 6.3.2).

9.1.2 Academic characteristics

From the 169 respondents, 39 reported not having a teaching qualification in mathematics. Of those, 33 stated having another undergraduate degree – the majority being in Education. This result shows that there are still some teachers teaching mathematics who have no first degree at all.

This was a surprising result because since 1996 having a degree in mathematics has been compulsory to teach mathematics in secondary schools and the government has recently invested a lot of effort and resources to encourage and allow all teachers who are teaching a subject that is not their speciality to gain an appropriate degree. Although surprising, it is understandable because, as in any part of the world, Brazil also suffers with a crisis in the teaching profession, especially in STEM-related subjects, as there are not enough teachers to teach all students.

Amongst teachers who stated that they held a teaching qualification in mathematics, 44 (35%) also stated that they held a second undergraduate degree[4]. Amongst these teachers, the majority had a second degree in Education (which is usually accompanied by a primary school teaching qualification), a teaching qualification in sciences or physics, or a degree in pure mathematics (without the teaching qualification).

Still in relation to their level of education, Figure 9.3 shows that most teachers (57%) had at least a postgraduate certificate. In this case, it also includes those teachers who stated not having the teaching qualification in mathematics.

[4]In Brazil, undergraduate degrees do not cover a second major. A second degree or a postgraduate degree is taken if one wants to gain a different speciality.

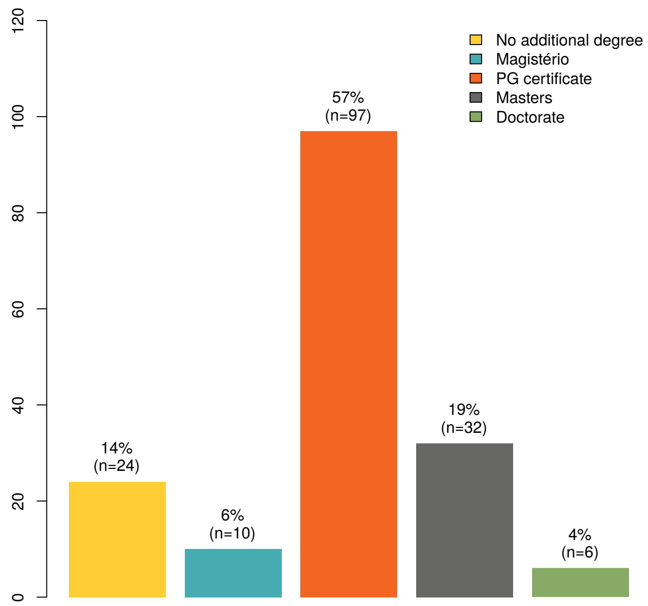


Figure 9.3: Highest level of education.

Table 9.1 better shows the distribution of respondents' degrees in relation to those who do or do not have a teaching qualification in mathematics.

Table 9.1: Degrees distribution in terms of having or not teaching qualification in mathematics.

Degrees	Teaching qualification in mathematics	
	Yes	No
No additional degree	21	3
Magistério	8	2
PG Certificate	71	26
Masters	25	7
Doctorate	5	1
Total	130	39

That is, Table 9.1 shows that 23% (39/169) of mathematics teachers do not hold a relevant teaching qualification. However, the majority (26/39) hold a postgraduate certificate, typically in a STEM-related subject. Occasionally it may be specifically in mathematics education (5/26), or in some other education-related area such as teaching in higher education, special needs education or management in education (7/26).

9.1.3 Employment characteristics

First of all, from the 169 participants, only 8 (5%) were not teaching at the time they answered the MaTFAKi. They stated that they were retired and/or in management positions (coordinators, supervisors or principal and vice-principals). From those who are currently teaching, the vast majority (95%) teach only in state schools and 5% in both (one shift in independent schools and the other in state schools).

In regards to teachers' years of experience, to facilitate the comparison, I have divided the previous 6-item scale into a 3-item scale: 'Early-career' teachers, with less than 10 years of experience; 'Experienced' teachers, between 11 and 20 years of experience; and 'Very-experienced' teachers with more than 20 years of experience in teaching. Figure 9.4 shows the distribution.

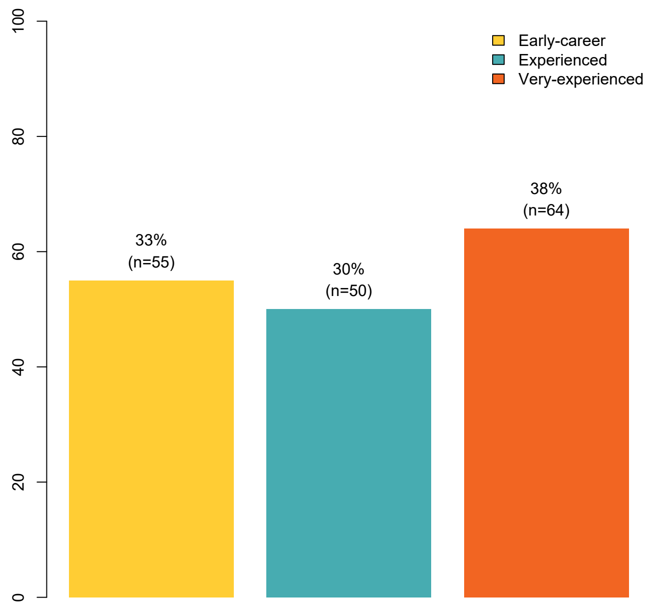


Figure 9.4: Years of experience.

As happened with the age groups, there was a good distribution between the different levels of 'Years of Experience' amongst the respondent teachers, with a higher proportion of very-experienced teachers (38%).

Although the 'Younger' group was larger than the other age groups, and therefore it was expected to have more early-career teachers, the cross-tabulation in Table 9.2 shows that 44% Middle-age teachers are very-experienced, which contributed to the higher proportion of this category.

Table 9.2: Cross-tabulation between Age and Years of Experience.

Age	Years of Experience			Row Total
	Early-career	Experienced	Very-experienced	
Younger	45 62%	28 38%	0 0%	73
Middle-age	7 16%	18 40%	20 44%	45
Older	3 6%	4 8%	44 86%	51
Column Total	55	50	64	169

In terms of the school level at which respondents were teaching at the time they answered the MaTFAKi, Figure 9.5 shows that there were teachers teaching all grades, with a higher proportion teaching Fundamental Education[5].

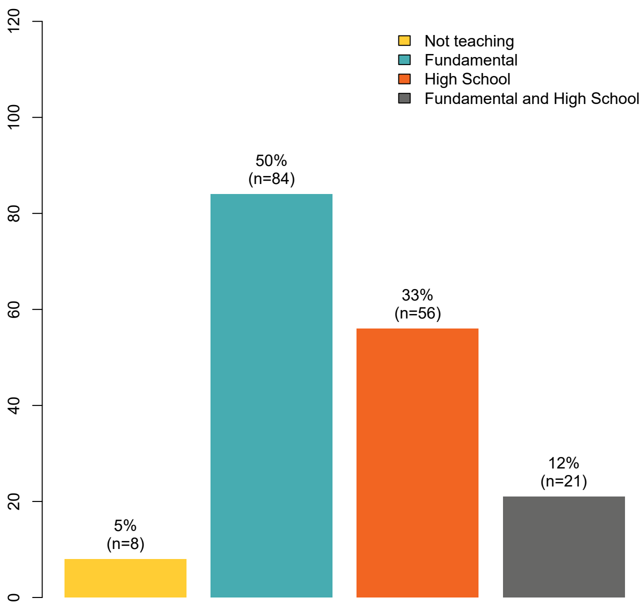


Figure 9.5: Grades being taught.

This result was also expected as there are more mathematics lessons at this level and hence a higher number of teachers. In addition, this level of education is compulsory, whereas Middle Education is not, which means more teachers are required due to the higher number of students.

[5] See Table 2.1 for reference.

9.2 Analysis of FD teachers' performance based on total scores

In this section, I present teachers' overall knowledge of FA, based on their total scores. As explained in Section 8.2.2, their performance will also be compared with the performance of teachers from the others states of Brazil (*BR*).

As explained in the previous chapter, for each question in the MaTFAKi, the four options were scored from 1 to 4 according to expert judgement of their adequacy in relation to the situation presented (with a score of 4 for the option considered the most appropriate). The sum of these scores, from all the MaTFAKi items, generated a total score for each teacher. As the minimum score of the options was 1 and the maximum 4, the total scores on the 20 items could range from 20 to 80. Figure 9.6 shows the distribution of FD respondents' total scores.

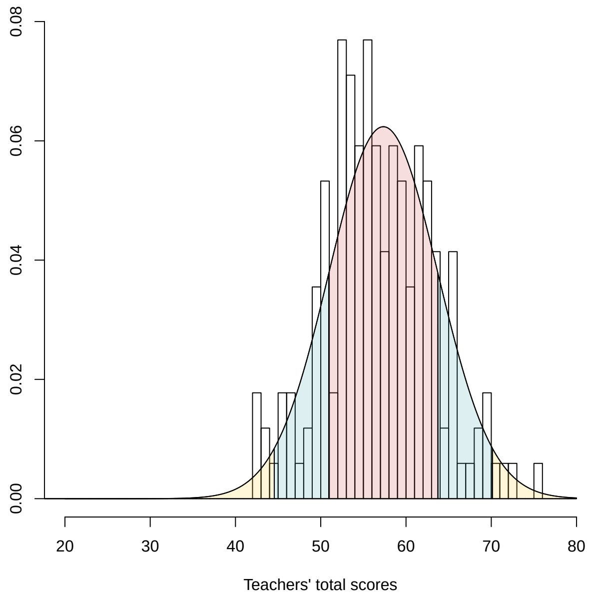


Figure 9.6: Histogram - FD responses.

The histogram shows that the total scores of the majority of teachers followed the 3-sigma rule, that is, they took values within three standard deviations of the mean ($mean = 57.34$, $sd = 6.40$). Looking at the total scores of individual teachers, they ranged from 43 to 76 which shows that no teacher had the maximum score of 80 but no teacher was close to the minimum score of 20.

Very roughly the 'average' response across teachers was at the level of the second most appropriate option (57.34 being close to 60), while the 'average' response of individual teachers ranged from, at the top of the distribution, the most appropriate option (76 being close to 80) to, at the bottom of the distribution, the less appropriate third (score 2) option (43 being close to 40).

The result of the Welch two sample t-test shows that when compared to the performance of *BR* teachers, the difference in means was statistically significant ($t = -4.4704, p < .0001$), and the 95% confidence interval of the difference in mean scores was between 1.69 and 4.34. That is, overall, the FD mathematics teachers surveyed had a lower performance when comparing solely their mean scores to those of *BR* teachers ($mean = 60.35$, $sd = 6.86$) who participated in the trial.

However, looking at the individual teachers (such as was done with the FD teachers above) the 'average' response across *BR* teachers was also at the level of the second most appropriate option (as 60.35 is also close to 60), while the 'average' response of individual teachers also ranged from, at the top of the distribution, the most appropriate option (75 being close to 80) to, at the bottom of the distribution, the less appropriate third (score 2) option (total score = 40).

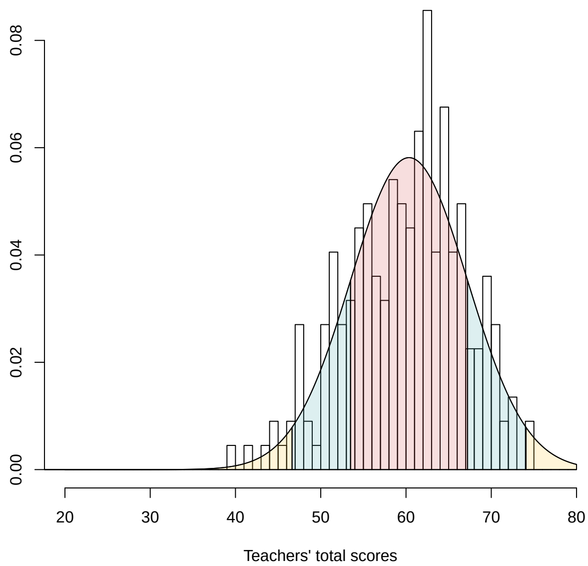


Figure 9.7: Histogram - *BR* responses.

Similarly, if the responses had been scored as a traditional multiple-choice test, with credit given for only one correct answer (the option scored 4), as in other pieces of research (e.g. Plake, 1993; Mertler & Campbell, 2005), the overall performance (both in the FD and *BR*) would have appeared much weaker than above, but the difference in their mean scores would still be statistically significant ($t = -3.8304$, $p < .001$).

In this case, the FD teachers' mean scores would be 8.17 ($sd = 2.82$) against a ceiling of 20. Thus, just over 40% of teacher responses were at the level of the most appropriate option. When compared to *BR* teachers, who had mean scores of 9.33 ($sd = 3.11$), their performance would still be inferior, as happened when the options were accorded scores 1–4. The 95% confidence interval of the difference in mean scores is between 0.56 and 1.76.

As explained in Section 8.2.2, to compare the difference in total scores not only based on their mean scores, a Mann Whitney U or Mann–Whitney–Wilcoxon (Hettmansperger & McKean, 2010; M. Hollander et al., 2013) was also performed. Both results of the Mann–Whitney–Wilcoxon test, from the 4 scores version ($U = 23848$, $p < .0001$) and the binary version ($U = 22957$, $p < .001$), were statistically significant and therefore it is possible to reject the null hypothesis and infer that there are differences between the two samples.

However, the only way to better understand the difference between the samples is by analysing the items separately. Similarly, this kind of total score analysis is not helpful to really understand which aspects of teachers' performance were better and in which aspects they could improve. Therefore, in the next sections, I will examine some particular items in terms of how FD teachers performed in that item compared to the other items and those in the *BR* sample.

9.3 Analysis of FD teachers' performance based on individual items

As explained above, in the next sections, I will present how FD mathematics teachers answered the items separately. Since the MaTFAKi comprised only 20 items, I believe that examination of each item, and of the resulting profiles, will offer a better holistic picture of teachers' answers than the overall scores presented above. Furthermore, analysis of each item is necessary for the purposes of this research, which is to provide diagnostic, actionable and formative information about teachers' knowledge of FA.

However, it is not the purpose of this analysis to fully comment on the specificities of each answer option, as this has already been done in detail in Chapter 7. That is, the knowledge teachers demonstrated has already been specified in the explanation of each option in the MaTFAKi.

To have a clear overview of how teachers' answers were distributed and how they were grouped in three performance categories (explained below), the percentages of each option for every question in the MaTFAKi can be found in Table 9.3. The colours represent those questions that were developed to assess different aspects of the same DoK.

Table 9.3: Percentage of teachers choosing each score option - FD results.

Category	Item_DoK	Score 1		Score 2		Score 3		Score 4	
Strengths	Q1_KIE	C	2%	D	6%	A	18%	B	74%
	Q5_KIE	D	4%	C	23%	A	13%	B	60%
	Q10_KIE	B	9%	D	12%	C	41%	A	37%
	Q18_KIE	C	18%	D	6%	A	18%	B	58%
	Q7_KHS	A	2%	D	24%	B	31%	C	43%
	Q14_KHS	C	5%	D	12%	B	20%	A	64%
	Q6_KLI	C	4%	B	8%	D	37%	A	50%
	Q20_KCL	C	11%	B	21%	D	25%	C	43%
	Q16_KCL	C	10%	A	24%	B	12%	D	54%
Weaknesses	Q3_KCL	D	17%	C	35%	A	28%	B	20%
	Q9_KCL	A	13%	C	46%	D	18%	B	23%
	Q11_KAM	D	30%	B	49%	C	9%	A	12%
	Q13_KAM	B	8%	A	55%	C	1%	D	37%
Divided performance	Q4_KAM	C	25%	D	22%	B	9%	A	43%
	Q2_KEF	B	36%	A	15%	D	8%	C	42%
	Q8_KEF	D	39%	B	8%	C	17%	A	36%
	Q15_KEF	A	15%	C	30%	D	21%	B	34%
	Q19_KEF	D	12%	A	37%	B	17%	C	34%
	Q12_KLI	B	24%	D	18%	A	40%	C	19%
	Q17_KLI	B	28%	D	10%	A	28%	C	34%

Table 9.3 shows that teachers' FA knowledge varied considerably across the four options that were presented. In some questions, it is clear that scores at the high end were the most favoured (e.g. Q10), but in others, the low end was favoured (e.g. Q11) or responses were spread fairly evenly over the four options (e.g. Q17).

It was possible to observe that there was no item in which an option was not chosen at all. The options that were least chosen were score 3 in Q13 (by 1% of teachers) and score 1 in Q1 and Q7 (by 2% of teachers). Equally, it is possible to observe that while score 4 was the most favoured in most questions (which I explore further in the next section), score 1 was also frequently chosen by a considerable number of participants. This result suggests that the MaTFAKi was generally well-designed to capture the range of FA knowledge amongst teachers.

Given that the overall ‘average’ performance was around score 3 (next-to-best option), one way of summarising the pattern for individual items was to examine the proportion of responses to each item which chose the best (score 4) or next-to-best option (score 3). Taking this approach, there were only three items (Q1, Q14 and Q6) in which this proportion exceeded 80%; six items (Q5, Q10, Q18, Q7, Q16 and Q20) in which the proportion exceeded 65%; and for further seven (Q2, Q4, Q8, Q15, Q19, Q12 and Q17) it was at least 50%. For the remaining four items (Q3, Q9, Q11 and Q13), the proportion choosing best or next-to-best options fell below 50%. Observing such differences in patterns of responses, the items were divided into three categories:

Strengths, encompassed areas in which teachers did relatively well. This was initially defined as those questions in which the option accorded score 4 was not just the most frequently chosen, but also chosen by most teachers. That is, those questions where more than 70% of respondents chose the answer considered to be the best in that situation. Two further items (Q16 and Q20) came very close to meeting this criterion, so the definition was relaxed to also include questions in which at least 65% of respondents chose an option that was considered to be the best or next-to-best in that situation.

Weaknesses, represented areas in which teachers did not perform so well. Here, questions that were included were those in which fewer than 50% of teachers scored 3 or 4. Those aspects in which their performance could not be considered as successful as the areas presented in the previous category.

Divided performance, included items with between 50 and 65% of responses at scores 3 and 4, and so between 35 and 50% at scores 1 and 2. Thus, for these items, teachers’ answers were distributed more across the spectrum of options, sometimes with high proportions at the two extremes. These are items that for some teachers would be classified as strengths and for others, weaknesses.

It was also possible to observe that for three DoK, all questions were included in the same performance category (KIE, KHS and KEF) and for the other three, the questions were included in two categories (KLI, KCL and KAM). Therefore, in the following section I will examine the DoK in which all items fell in the same performance category. In Section 9.3.2 I will analyse the DoK in which the items fell in more than one category. At this stage, only the results of the FD teachers will be explored.

9.3.1 Items grouped in the same performance category

As explained above, there were three DoK for which the items presented a similar distribution of responses so that they ended up being included in the same category of analysis. In two of these DoK, FD teachers' performance was better and therefore considered as their 'strengths' (KIE and KHS), and in the third (KEF), teachers had a 'divided performance' (see Table 9.3).

In relation to teachers' KIE, it was possible to observe that for three of the four items (Q1, Q5 and Q18), the majority of teachers chose score 4 (74%, 60% and 58% respectively). Q1 also deserves attention as being the only item in the whole scale in which more than 70% of teachers identified the option considered to be the best in the situation presented.

In Q10, where the majority did not choose score 4 (37%), score 3 (next-to-best) was most chosen (41%) and in this case, this option also represents a good interpretation of students' thinking and therefore of their learning, which is aligned with the results of the other three items assessing this DoK.

In relation to teachers' KHS, although it was assessed by two items only (Q7 and Q14), both fell in the 'strengths' category and both followed the 'ideal' pattern. Most teachers choosing score 4 (43% and 64% respectively) followed by score 3 (31% and 20%), then score 2 (24% and 12%) and just a few choosing score 1 (2% and 5%). The results of both items show that FD teachers performed well in terms of helping students to use assessment information for the improvement of their learning.

Finally, in relation to the KEF, FD teachers presented a divided performance in all four items assessing this DoK. In Q2 and Q8, their performance was divided between scores 4 and 1; whereas in Q15 and Q19, between scores 4 and 2. In all questions, these options represent opposite ideas.

In Q2, while the 42% of teachers that chose the score 4 option identified the best type of feedback to give to students in the situation presented, the 36% of teachers that chose score 1 judged that the best strategy would be not to give any feedback and wait for students to ask for the teacher's help.

In Q8, both options finish by returning the question to the student. The difference is that for the 39% of teachers who chose the score 1 option, this question should be very specific and tell students their mistake and therefore they know straight away what to do (*can you accept the negative values?*); whereas for the 36% who chose the score 4 option, the question prompts students to look for their mistake (*can you accept all the values you found?*).

This result is aligned to the results of Q15 and Q19. In Q15, for the 34% of teachers who chose the score 4 option, it is better to give students some direction about what they missed, but not do the work for them (*Can you think of what else you need to do to find one of the answers provided?*). For the 30% who chose the score 2 option, it is better to specifically tell students what they need to do (*Now you need to rationalise the denominator so you can find one of the answers provided*).

In Q19 the 37% of teachers who chose the score 2 option believed that the best strategy, in an example of oral feedback, would be to tell the student the result of his thinking straight away: *you will get 21% of 50 which is not what the problem is telling us*. On the other hand, for the 34% of teachers that chose the score 4 option, the best strategy would be to ask the student to explain how he came up with that result (*Can you show me how you came up with 0.21?*) and then encourage his thinking with a follow up prompt: *Tell me why you think this may work in this case.*

Finally, within the items assessing teachers' KEF, Q8 also deserves attention as being the only item in the whole scale for which score 1 was most favoured.

9.3.2 Items separated in different performance categories

In this section, I explore the items which, although developed to assess the same DoK, presented different patterns of responses and therefore fell in different performance categories. This was the case with KCL, KLI and KAM items.

Regarding this division, the most interesting aspect lies in how teachers answered the questions assessing the KCL, as two of them were included in the 'strengths' category and two in the 'weaknesses'. That is, in two of them teachers performed relatively well, and in the other two, their performance could not be considered so good. Looking at the characteristics of Q3 and Q9, which were included in the weaknesses category, it is possible to observe that in both questions, score 2 (which was the option most favoured, 35 and 46% respectively) relates to the teacher solving the problem or showing the correct solution on the board.

In the case of Q16 and Q20, both are related to teachers giving instructions to students about what to do next. That is, they both ask teachers to judge the best way of advising students of what to do after feedback has been given, and therefore it was expected that teachers would answer the questions similarly.

This division was also observed in relation to questions assessing KLI, with two items in the ‘divided performance’ category (Q12 and Q17) and one in the ‘strengths’ (Q6). However, because each question assessed a different aspect of KLI, this division might have happened because of particularities of each of these aspects.

In Q6, teachers were required to analyse what went wrong when presenting the learning intentions to students and what would be the best way to deal with them. 50% of teachers chose the score 4 option, and therefore recognise that it is important to use some strategy to check if students understood the learning intentions.

In Q12, which refers to teacher checking how the students understood the learning intention, 40% of teachers chose score 3, and therefore had a relatively good performance. However, the second most chosen option is represented by score 1, with 24% of teachers believing that the best strategy would be to write the learning intentions on the board so students *can identify them in the exercises and the feedbacks*.

In Q17, teachers were presented with a situation in which they had to judge the best way of devising success criteria. In that case, there was also almost an equal division among scores 1, 3 and 4. Because this is the only question which deals with the success criteria, and therefore has very specific characteristics, it is not possible to compare its contents with Q12 which was included in the same category, or with any other question. Therefore, the interpretation of how FD mathematics teachers understand this important aspect needs to be based only on their answers to Q17 itself and/or a comparison of their results with the results of *BR* teachers (see Section 9.4).

However, comparing the two extremes (options representing scores 4 and 1, respectively), 34% of teachers believe that the criteria should be devised together with students and 28% of teachers believe the criteria should be defined by the teacher and explained to students, following a top-down, teacher-centred approach.

In fact, in the four items discussed above (Q3, Q9, Q12 and Q17), in which options in the low end of scores attracted teachers, and therefore led to their performance being considered not so good, all refer to a top-down, teacher-centred approach: the teacher defining the criteria for students to follow, teachers showing the correct solution on the board, etc.

Finally, in relation to teachers' KAM, two items fell in the 'weaknesses' category (Q11 and Q13) and in one they were considered to have a 'divided performance' (Q4).

For both items considered as teachers' weaknesses, FD teachers have demonstrated a misconception in terms of the purpose and potential of assessment methods.

In Q13, 55% of teachers believe that the purpose of the 'continuous recovery' is *to give the students different opportunities to recover the mark along the term*, instead of focussing on their learning.

In Q11, 49% of FD teachers do not identify the potential of multiple-choice questions for eliciting evidence of students' learning.

In Q4, which was the item assessing their KAM that fell under 'divided performance', 43% of teachers chose the score 4 option and therefore recognise that the activities chosen to assess students were not in accordance with the learning intentions.

Analysing the results of Q4, it is possible to say that this item could also be considered as one of the teachers 'weakness' due to the content of the options. That is, option A ($score = 4$) can be compared to the other three options together because option A is the only one in which the respondent teacher identifies the assessment method was not appropriate for the teacher's intention. Therefore, taking the three other options together, 57% of teachers actually do not identify the problem with the activity.

Within teachers' KAM, Q11 also deserves attention as being the item in the whole scale which teachers' performance was the lowest, with only 12% choosing the best option, and 79% of teachers choosing scores 1 and 2.

9.4 Analysis of FD teachers' performance compared to the *BR* sample

In this section, I explore how FD teachers' answers differ from those of *BR* teachers. First I present Table 9.4 with the percentages of *BR* teachers' answers, using the same rationale to divide their answers into 'strengths', 'weaknesses' and 'divided performance'. Next, I present the differences in percentage between the FD and *BR* samples (Table 9.5).

Table 9.4: Percentage of teachers choosing each score option - *BR* results.

Category	Item_DoK	Score 1	Score 2	Score 3	Score 4
Strengths	Q1_KIE	3%	8%	10%	79%
	Q5_KIE	5%	29%	9%	57%
	Q10_KIE	7%	11%	44%	38%
	Q18_KIE	19%	6%	14%	60%
	Q7_KHS	6%	22%	26%	46%
	Q14_KHS	10%	10%	23%	47%
	Q16_KCL	11%	20%	9%	60%
	Q20_KCL	8%	27%	25%	41%
	Q6_KLI	9%	13%	26%	52%
	Q17_KLI	18%	15%	27%	40%
	Q2_KEF	14%	18%	17%	50%
	Q8_KEF	27%	8%	26%	39%
	Q15_KEF	13%	17%	19%	51%
	Q4_KAM	16%	18%	11%	55%
	Q11_KAM	7%	8%	47%	38%
Weaknesses	Q13_KAM	9%	51%	2%	38%
Divided performance	Q3_KCL	14%	30%	26%	29%
	Q9_KCL	17%	28%	23%	32%
	Q12_KLI	15%	31%	17%	37%
	Q19_KEF	21%	29%	18%	33%

Looking at Tables 9.3 and 9.4, two main characteristics should be highlighted.

First is in relation to how the items were divided into the performance categories. In the FD, questions representing the same DoK were better grouped, whereas in the *BR* this grouping was possible only with questions representing examples of KIE and KHS (featuring as teachers' strengths in both samples).

Second is that the *BR* sample presented considerably more aspects in the 'strengths' category, and fewer aspects in the 'weaknesses'. However, although in different proportions, both regions had more aspects that could be considered as teachers' strengths (9 items in the FD, 15 in the *BR*), followed by aspects in which they presented a divided performance (7 and 4 items respectively) and fewer aspects which were considered weaknesses (4 and 1 items respectively).

In any case, it was possible to notice that some questions overlap, i.e., they were included in the same categories in both the FD and *BR* and therefore teachers from both samples presented similar performance:

- Strengths: Q1, Q5, Q6, Q7, Q10, Q14, Q16, Q18, Q20
- Weaknesses: Q13

- Divided views: Q12, Q19

Yet, even when an aspect was considered to be one of the FD teachers' strengths, only in three of the questions that overlap, was the proportion of teachers choosing score 4 higher in the FD than the *BR*. Therefore, if the analysis of FD teachers' strengths was based solely on those questions in which they did better than *BR* teachers, that would be the case only for Q5, Q14 and Q20 (and in none of them the difference is statistically significant, see Table 9.5).

For a better view of how their answers differed in each option of each item, Table 9.5 presents the differences in percentages between the FD and *BR* samples.

Table 9.5: Differences in responses between FD and *BR* samples.

Item_DoK	Score 1	Score 2	Score 3	Score 4
Q1_KIE	-1%	-2%	**8%**	-5%
Q5_KIE	-1%	-6%	4%	3%
Q10_KIE	2%	1%	-3%	-1%
Q18_KIE	-1%	0%	4%	-2%
Q2_KEF	**22%**	-3%	**-9%**	-8%
Q8_KEF	**12%**	0%	**-9%**	-3%
Q15_KEF	2%	**13%**	2%	**-17%**
Q19_KEF	**-9%**	8%	-1%	1%
Q3_KCL	3%	5%	2%	-9%
Q9_KCL	-4%	**18%**	-5%	-9%
Q16_KCL	-1%	4%	3%	-6%
Q20_KCL	3%	-6%	0%	2%
Q4_KAM	**9%**	4%	-2%	**-12%**
Q11_KAM	**23%**	**41%**	**-38%**	**-26%**
Q13_KAM	-1%	4%	-1%	-1%
Q6_KLI	-5%	-5%	**11%**	-2%
Q12_KLI	9%	**-13%**	**23%**	**-18%**
Q17_KLI	**10%**	-5%	1%	-6%
Q7_KHS	-4%	2%	5%	-3%
Q14_KHS	-5%	2%	-3%	7%

The numbers in bold are the ones in which the results of the 'prop.test'[6] were statistically significant. However, because the numbers are large, relatively small effect sizes may prove statistically significant, and therefore, considering that there is a danger in discussing the results focussing on significance level at the expense of effect size, I will discount any item where the difference is less than 10%.

[6]See Section 8.2.2 for explanation.

With that being said, the item that deserves the most attention is Q11, as it is the only one in which the difference in proportion was observed in **all four** score options ($FD_1 = 30\%, BR_1 = 7\%, \chi_1^2 = 34.45; FD_2 = 49\%, BR_2 = 8\%, \chi_1^2 = 80.22; FD_3 = 9\%,$ $BR_3 = 47\%, \chi_1^2 = 61.28; FD_4 = 12\%, BR_4 = 38\%, \chi_1^2 = 31.18, p < .0001$). While this item was considered one of the FD teachers' weaknesses, in the *BR* it featured as one of their strengths. The performance of FD teachers in this item, was much lower than that of *BR* teachers.

In Q12, the difference in proportions was observed in **three** of the four options, with a higher percentage of FD teachers choosing score 3 ($FD_3 = 40\%$, $BR_3 = 17\%$, $\chi_1^2 = 24.79, p < .0001$) and a lower percentage choosing scores 2 ($FD_2 = 18\%$, $BR_2 = 31\%$, $\chi_1^2 = 7.81, p < .01$) and 4 ($FD_4 = 37\%$, $BR_4 = 19\%$, $\chi_1^2 = 14.86, p < .001$). Although the proportions were different, Q12 was included in the 'divided performance', both in the FD and *BR*. In this case, however, there was a higher proportion of teachers in the *BR* that identified the best option in the situation presented. For some reason, FD teachers were more attracted to the score 3 option.

In Q15, the difference was observed in **two** options (scores 2 and 4). The performance of FD teachers in this item was also lower than that of *BR* teachers, both because of the higher proportion of score 2 ($FD_2 = 30\%$, $BR_2 = 18\%$, $\chi_1^2 = 8.53, p < .01$) and the lower proportion of score 4 ($FD_4 = 34\%$, $BR_4 = 51\%$, $\chi_1^2 = 11.41, p < .001$).

Finally, there were six items in which the difference in proportions was observed in **only one** of the four options.

In Q2, Q8 and Q17, there was a higher proportion of FD teachers choosing score 1 ($FD_1 = 36\%$, $BR_1 = 14\%$, $\chi_1^2 = 24.96, p < .001$; $FD_1 = 39\%$, $BR_1 = 27\%$, $\chi_1^2 = 6.31, p < .05$; $FD_1 = 28\%$, $BR_1 = 18\%$, $\chi_1^2 = 4.2809, p < .05$, respectively), and therefore their performance in these items could be considered lower than that of *BR* teachers.

In Q9, the difference was observed in score 2 ($FD_2 = 46\%$, $BR_2 = 28\%$, $\chi_1^2 = 12.39, p < .001$) and in Q4, in score 4 ($FD_4 = 43\%$, $BR_4 = 55\%$, $\chi_1^2 = 8.24, p < .05$). In both cases, the performance in the *BR* was better than that of the FD sample. While Q9 was considered one of the FD teachers' 'weaknesses', *BR* teachers had a 'divided performance'. Q4, on the other hand, was considered one of the FD teachers' 'weaknesses', and one of the *BR* teachers' 'strengths'.

In Q6, this difference was observed in score 3 ($FD_3 = 37\%$, $BR_3 = 26\%$, $\chi_1^2 = 5.0747, p < .05$, respectively), with a higher proportion of FD teachers.

Overall, it is also possible to notice that of the items which presented statistically significant differences in proportions in scores 1, 2 and 3 the FD sample had a higher proportion of teachers favouring this option. Looking at score 4, on the other hand, the percentage of *BR* teachers was higher in all of them.

Additionally, in three of the four items developed to assess different aspects of the KEF, it was possible to observe the 10% difference in proportions in at least one of the options. Similarly, no difference above 10% was observed in the two items assessing the KIE or KHS (considered strengths in both samples).

The results presented above, therefore, shows the *BR* sample consistently performing higher than the FD sample.

In the sections above, I have presented how FD teachers answered each question and how I divided them into three categories – their strengths, weaknesses and those aspects in which they presented a divided performance; and how the items could be grouped based on the responses behaviour. I also compared the performance of FD teachers to the performance of *BR* teachers. In the next section, I present some results in regards to how some answers varied in terms of FD teachers' background information.

9.5 Analysis of FD teachers' performance based on their background information

This section aims to use the background information available about the FD sample to answer the second set of research questions, exploring the items for which there was a statistically significant variation in teachers' knowledge based on their academic and employment characteristics[7].

9.5.1 In terms of years of experience

In terms of 'Years of Experience', there were generally no differences between the three experience groups. The only exceptions were Q16 and Q17.

The analysis of Q16 is shown in Table 9.6.

[7] As explained in Section 8.2.2, the analysis of the significance and association was made through a Goodman Kruskal's gamma test (Field et al., 2012; Key, 1997) and the results are shown as ($\gamma = q, p < \alpha$), where q is the observed value of the γ statistic and p is the p-value, which is compared to the significance level α.

Table 9.6: Cross-tabulation between Q16 and Years of Experience.

Q16	Years of Experience			Total
	Early-career	Experienced	Very-experienced	
Score 1 Count % within Experience	4 7%	6 12%	7 11%	17 10%
Score 2 Count % within Experience	22 40%	10 20%	8 12%	40 24%
Score 3 Count % within Experience	5 9%	7 14%	8 12%	20 12%
Score 4 Count % within Experience	24 44%	27 54%	41 64%	92 54%
Total	55	50	64	169
($\gamma = 0.224, p < .05$)				

Overall, most teachers (54%) chose score 4, but this proportion rises between 'Early career' (44%), 'Experienced' (54%) and 'Very experienced' groups (64%). The most popular of the other options was that given score 2 (by 24% of teachers overall), and here the proportion falls between 'Early career' (40%), 'Experienced' (20%) and 'Very experienced' groups (12%). Around 10% of teachers at each level of experience choose each of the score 1 and 3 options. Thus the proportion of teachers choosing the best option rises with experience, while the proportion choosing the next-to-worst option falls.

The second item for which it was possible to find statistically significant variation was Q17. Table 9.7 shows how teachers' answers were distributed by years of experience.

Table 9.7: Cross-tabulation between Q17 and Years of Experience.

Q17	Years of Experience			Total
	Early-career	Experienced	Very-experienced	
Score 1 Count % within Experience	25 46%	12 24%	10 16%	47 28%
Score 2 Count % within Experience	2 4%	6 12%	9 14%	17 10%
Score 3 Count % within Experience	11 20%	14 28%	23 36%	48 28%
Score 4 Count % within Experience	17 30%	18 36%	22 34%	57 34%
Total	55	50	64	169
($\gamma = 0.192, p < .05$)				

Around 30% of teachers at each level of experience chose score 4. However, the proportion of score 3 rises between 'Early career' (20%), 'Experienced' (28%) and 'Very experienced' groups (36%) while the proportion of score 1 falls considerably between 'Early career' (46%), 'Experienced' (24%) and 'Very experienced' groups (16%).

In conclusion, therefore, the most striking finding is that there appears to be no statistically significant 'Experience' effect with 18 of the 20 items.

In the next section, the same analysis will be done in terms of 'education'.

9.5.2 In terms of education

I divided the variations in terms of education in two parts: *1*) in terms of having or not the teaching qualification in mathematics, and *2*) in terms of additional degrees.

9.5.2.1 In terms of having or not teaching qualification in mathematics

In general there was no difference in performance on items according to whether or not teachers held a teaching qualification. However, there were two aspects in which we could see a statistically significant difference between teachers with or without the teaching qualification in mathematics.

The first aspect was in regards to Q11. As Table 9.8 shows, the majority of teachers, independently of whether they have the relevant qualification or not, are concentrated in scores 1 and 2. The main difference between those teachers with and without a teaching qualification is that the latter gave fewer score 1 responses and more score 3 responses, presenting a stronger profile of performance.

Table 9.8: Cross-tabulation between Q11 and Teaching Qualification.

Q11	Teaching qualification		Total
	Yes	No	
Score 1 Count	43	7	50
% within Qualification	33%	18%	30%
Score 2 Count	63	19	82
% within Qualification	48%	49%	48%
Score 3 Count	8	8	16
% within Qualification	6%	21%	9%
Score 4 Count	16	5	21
% within Qualification	12%	13%	12%
Total	130	39	169
($\gamma = 0.304, p < .05$)			

The second aspect in which there was a statistically significant difference among teachers' answers was in relation to Q20 (Table 9.9).

Table 9.9: Cross-tabulation between Q20 and Teaching Qualification.

Q20	Teaching qualification		Total
	Yes	No	
Score 1 Count	13	5	18
% within Qualification	10%	13%	11%
Score 2 Count	21	14	35
% within Qualification	16%	36%	21%
Score 3 Count	35	8	43
% within Qualification	27%	20%	25%
Score 4 Count	61	12	73
% within Qualification	47%	31%	43%
Total	130	39	169
($\gamma = 0.307, p < .05$)			

Table 9.9 shows that respondents without the teaching qualification in mathematics were more likely to choose score 2 and less likely to choose scores 3 and 4, a weaker pattern of response than those with a teaching qualification.

Again, in conclusion, therefore, the result is that there appears to be no statistically significant 'Teaching Qualification' effect with 18 of the 20 items.

9.5.2.2 In terms of additional degrees

There was no variation that was statistically significant when comparing FD teachers' answers with their level of education. I believe detecting anything other than very gross differences would be difficult within this dataset because there are some categories (e.g. doctorate) that had almost no representation.

9.6 Summary

In this chapter, I began by presenting participants' characteristics and argued that although the total scores of FD mathematics teachers were below the scores of *BR* teachers, a deeper analysis of each aspect was necessary to better understand their FA knowledge. I then went on to suggest that there are indeed some aspects in which teachers did well and some in which they need to improve, but this variation is not particularly dependent on their years of experience or whether they do or do not hold a teaching qualification in mathematics. Other insights that emerged from the data were also presented.

A summary of these findings and of the principal issues and suggestions which have arisen in this study are provided in the next chapter, in which I aim to discuss the answers to the research questions which inspired this inquiry.

CHAPTER 10

DISCUSSION

The previous chapter presented the findings of the study in two parts: *1)* in terms of FD teachers' performance, not only based on how they answered each question of the MaTFAKi, but also comparing their results to those of teachers in the other states of Brazil; and *2)* based on how each question behaved and how it could be analysed comparing its options and content. In this chapter, the aim is to discuss these results. The focus is on better understanding teachers' knowledge to be able to feed back to teachers themselves, and those working with them; but also to analyse the MaTFAKi as a new instrument to assess mathematics teachers' knowledge of FA. Therefore, the discussion of the findings is also divided in this way.

Section 10.1 evaluates the data and subsequent analysis to establish the generality and limitation of results. Section 10.2 offers answers for the first research question, by discussing the findings with regard to FA theory, mathematics teaching and learning, and their implications for educational practice. Section 10.3 provides answers for the second set of research questions exploring the systematic patterns of variation in teachers' FA knowledge. Section 10.4 discusses the MaTFAKi as a new instrument for assessing mathematics teachers' FA knowledge, in an attempt to answer the final research question.

10.1 Evaluating the data

To be able to confidently discuss the results of the study, the evaluation of the appropriateness of the data is inevitable. Without a carefully drawn sample, an adequately designed research instrument, and satisfactorily processed data, statistical results cannot be supported and generalised.

As discussed in Chapter 9, the sampling strategy was developed in such a way that all teachers would have an equal chance to answer the MaTFAKi. The MaTFAKi was sent to all FD secondary schools, and other measures were taken to guarantee that all emails were received. In the final dataset, although it could be considered a relatively small sample, there was adequate representation in the data as the characteristics of teachers were well divided in terms of gender, age and years of experience.

All items were repeatedly piloted and evaluated to suit the context in which teachers were teaching, and to make sure that each question was assessing what it was supposed to assess, ensuring the validity of the MaTFAKi. The data was adequately prepared to meet the assumptions of each statistical test used and findings were reported by acknowledging any inherent limitations in the data (e.g. the unsuitability of mean scores).

Thus, the careful development of the MaTFAKi, the different pilot phases and the treatment of the data provided solid backing for the findings of this study. Nevertheless, I will remain cautious about generalising the results. In the upcoming sections, I will offer some explanations, conjectures, and hypotheses about results that could and should be further developed and confirmed in future studies.

10.2 Answering the first research question

> What do secondary mathematics teachers in the Brazilian FD know about FA in general and the idea of feedback in particular, as measured by the MaTFAKi?

The discussion of the first research question is divided into two parts. Section 10.2.1 approaches FA in general, combining teachers' answers to various questions in the MaTFAKi. Section 10.2.2 approaches the idea of feedback in particular, based on teachers' answers to those questions which were developed to assess this aspect of their FA knowledge.

10.2.1 Knowledge of formative assessment in general

As discussed in Chapter 3 and as the results of the statistical analysis have shown, FA has a multidimensional nature and therefore involves a lot of interlinked aspects.

With the MaTFAKi, some snapshots were assessed and in this section my purpose is to discuss them together to answer the first part of the first research question.

Even though I will sometimes mention a specific DoK separately, the purpose will be to bring together those aspects of FA that were possible to observe based on the answers across questions. The discussion will focus on the effects and implications of teachers' answers on both FA and mathematics teaching and learning. Some Brazilian cultural aspects and my own experience will also be used to interpret some results. I will start discussing FD teachers' overall level of FA through the results of their total scores.

10.2.1.1 Teachers' average total scores

The result of the teachers' performance in the MaTFAKi as a whole was presented in two different ways: *1)* considering OMC responses (with different scores for each option), and *2)* considering the answers as traditional multiple-choice responses, with only one correct answer. The results were presented in two different ways to show that analysing teachers' performance only based on scores does not tell much about their knowledge and can lead to superficial analysis.

When using OMC responses, teachers' average total scores were above average ($mean = 57.34$, $sd = 6.40$) with the responses across teachers at the level of score 3 option. This result gives the impression that FD secondary mathematics teachers have an acceptable knowledge of FA as score 3 was considered the next-to-best answer. Even if the *BR* teachers showed relatively higher average scores ($mean = 60.35$, $sd = 6.86$), both are at the level of score 3 option and could be considered as acceptable results.

On the other hand, when analysing teachers' scores using the scoring convention of a traditional multiple-choice questionnaire, the mean scores ($mean = 8.17$ $sd = 2.82$) were below average or near 41% correct, and below the *BR* results ($mean = 9.33$ $sd = 3.11$). Given that in Brazil usually the passing score on classroom tests is 50%, most teachers participating in the MaTFAKi would receive a failing grade based on their demonstrated knowledge of optimal FA strategies.

These latter results, when compared to other studies using traditional multiple-choice questionnaires (Campbell et al., 2002; Mertler, 2003; Mertler & Campbell, 2005; Plake & Impara, 1997), suggest that FD teachers had a lower performance than other teachers in comparable contexts.

Although the aforementioned studies were not conducted only with mathematics teachers and focused on teachers' assessment competencies based on the Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, & NEA, 1990), teachers' scores were always above average – from 60 (Campbell et al., 2002) to 66% correct (Plake & Impara, 1997). At the same time, it is important to acknowledge that the comparison between these studies depends on the relative degree of difficulty of the different instruments. It may be the case that the MaTFAKi was harder than the others.

These studies solely measure teachers' overall performance and/or show whether teachers do or do not know the specific aspect being assessed by an item. That is, if the teacher did not choose the correct answer, it is not possible to assess partial knowledge (Ben-Simon et al., 1997). With the MaTFAKi, it is possible to identify what teachers do not know, but also to understand what aspects they already know about that topic, based on the option that they chose. Therefore, analysing only their total scores, using either the ordered or the traditional multiple-choice design, is a very simplistic way of analysing teachers' knowledge, even when trying to understand their FA knowledge in general.

Likewise, considering that the MaTFAKi was developed in a OMC response format, presenting teachers' total scores as 'above average' could indicate that teachers are doing well and that there is no need to take any actions in regards to their knowledge of FA. In other words, although it looks as if the average response by teachers was at the level of score 3 option, the results presented in the previous chapter have shown that, in fact, there were aspects where teachers' performance was relatively lower than the average scores suggest; and others where teachers were able to identify the score 4 option and therefore exceeded the average.

Following this rationale, in the next section I will discuss these areas and how these are related to and impact upon the teaching and learning of mathematics.

10.2.1.2 Exploring different areas of teachers' knowledge

In this section, the aim is to discuss the first part of the research question – what teachers know about FA in general – exploring their answers to various examples of FA included in the MaTFAKi.

First of all, teachers seem to know how to interpret the evidence of students learning, as their performance was relatively high in all examples assessing their KIE. In these examples, the respondents were presented with students' mathematical answers and were required to interpret them. However, the interpretation that was available in the four answer options varied from question to question based on the scenario in which they were included. In some, teachers were solely presented with a student mathematical solution (Q1 and Q18) and in others, there were more elements included – a conversation among students (Q5) or a situation of peer-assessment (Q10). Regardless, in three of the four examples, most teachers were able to identify the best answer. The exception was Q10.

The results were somehow not unexpected and could be explained by the fact that previous studies (Lysaght & O'Leary, 2013) and particularly those with Brazilian mathematics teachers (Albuquerque, 2012; Camargo & Ruthven, 2014; Camargo, 2015) have shown that peer-assessment (referred to only in Q10) is not a common practice among them. Therefore, teachers might not recognise its value in supporting students' improvement, resulting in it not being included in teachers' interpretation of their learning. In this case, score 3, which was the most chosen, included the interpretation of the mathematical solution only, which is aligned with the results of the other examples of KIE.

Considering that the ability to interpret evidence of students' learning is mostly dependent on teachers' content knowledge (Hodgen & Wiliam, 2006; Son, 2013) and on their knowledge of how the learning of mathematics occurs (Ball et al., 2008; Brookhart, 2011), it could be argued that participants have well-developed knowledge of mathematics and students' mathematical thinking and therefore they performed well on these examples due to this knowledge and not to particular knowledge related to FA.

Even so, considering that the overall idea of FA is about using assessment information to plan the next steps (ARG, 2002; Black & Wiliam, 1998), "strong content knowledge and understanding of learning progressions are likely precursor skills to using assessment information accurately" (Schneider & Meyer, 2012, p. 21). Therefore, it is important for teachers to have well-developed knowledge of mathematics and of students' mathematical thinking to be able to interpret learning, inform the next steps for teaching and learning, provide effective feedback, and help students to improve (Furtak, 2012; Gottheiner & Siegel, 2012; Ní Chróinín & Cosgrave, 2013; Yin, Tomita, & Shavelson, 2014).

This is not to say that content knowledge is enough for teachers to be able to effectively put FA into practice (Black et al., 2003), to understand students mathematical thoughts (Ball et al., 2008) or even to use the interpreted information to plan the next steps (Good, 2011). Identifying errors does not help to determine what to do about them. Research on sixth-grade teachers' ability shows that using assessment information to plan the next step of instruction tends to be the most difficult step for teachers (Heritage, 2007; Heritage, Kim, Vendlinski, & Herman, 2009).

On the other hand, M. Askew (2008) and Millett, Askew, and Brown (2004), concluded that teachers who had a sound subject knowledge base were more effective at interpreting student mathematical understanding and making good use of it. This included decision making during classroom interactions and planning the next lesson, such as, particular content that students would be taught, activities students would carry out, and how students would engage with the content. Although the MaTFAKi did not include examples of all these situations, the results suggested that participants have good subject knowledge on the examples included, which could be considered as the first step towards being able to adjust instruction. In this regard, the examples assessing teachers' KHS and KCL provided some information.

In saying so, helping students use assessment information (KHS) is another aspect of FA that FD mathematics teachers seem to know. The examples in the MaTFAKi included teachers judging how best to intervene in three students' discussion to help them analyse their approach to problem solving (Q7), and how to advise students on how to use the results of some quizzes to improve their learning (Q14). Although there were only two examples assessing teachers' KHS, and they assessed two different aspects, teachers performed well in both.

In regards to intervening in the students' conversation, the results suggested that teachers understand that it is not necessary to give students the answer straight away to help them to use the information (Brookhart, 2008; Wiliam, 1999), but instead to guide them on how to get to the expected answer (Q7.C). At the same time, the results also led to the belief that FD mathematics teachers recognise strategies of problem solving (Michalewicz & Fogel, 2013; Polya, 2014) and are taking them in to account to help students evaluate whether the solution is a reasonable response to the problem.

Even though there was also a considerable percentage of teachers that would direct students to analyse *whether divide by a half and divide in half is the same thing* (Q7.B), in both situations it appears that teachers know it is important to offer some kind of suggestion to help students identify the mathematical concept that they have to understand to solve the problem.

These results could also be compared with teachers' divided performance in terms of the characteristics of effective feedback, as discussed in Section 10.2.2. It seems that some teachers tend to believe that the best way to respond to students is by giving them a more direct pathway to follow, whereas others provide more of a guidance so students can figure out 'Where to next?' (Hattie & Timperley, 2007), which is more likely to lead to actual learning (Brookhart, 2008; Wiliam, 1999, 2011).

In terms of helping students use the results of quizzes taken over the course of the term, teachers' answers suggest that they know how to guide students to look back to the quizzes and identify the areas they struggled with. It looks as if teachers understand that in that situation they are providing information to help students to progress (Earl & Katz, 2008), instead of making them focus on areas they have already mastered (Q14.A).

Though this could be seen as a 'teaching to the test' strategy, the results from Black et al.'s study (2003) have shown that this is a recognised strategy for teachers when using summative assessments with formative purposes as it helps "to achieve a more positive relationship between the two" (p. 55) and engage students in a reflective review of the work they have done to enable them to plan their revision effectively. In Lysaght and O'Leary's study (2013), however, teachers reported to be more common practice in their lessons for the teacher themselves to use teacher-made assessment for diagnostic purposes than for the students to use them to review their own work.

Still referring to teachers advising students on how to use quizzes as feedback, the second most chosen option was the one in which the teacher suggests students look for help (Q14.B), which could indicate that the teachers who chose this option share the views of the teachers in Lysaght and O'Leary's study (2013) or are still not comfortable with the idea of students as autonomous learners (O'Shea, 2015).

In fact, teachers' answers to other questions in the MaTFAKi illustrated that this may be the view of many FD mathematics teacher, as a considerable number of them judged the best option to be the one in which the teacher would move the next step to the board, correct the exercise, or ask a student to show the correct solution to the whole group. Their focus on showing the correct answer, or stating exactly what to do, is not aligned to the principle of encouraging students to be the owners of their learning (Wiliam & Leahy, 2015). This necessity of taking control, showing the correct solution, or specifically telling students how to get the correct answer, might be a reflection of teachers' mathematical knowledge and their view of how mathematics should be taught.

Telling students what to do, or giving them advice on how to act, is not inherently a bad thing. However, from analysing the results of various questions it seems that respondents need to find a better balance between their own, and students' participation in the teaching and learning process, depending on the situation they are in. For example, looking at the characteristics of Q3 and Q9, it is possible to observe that the answer options that were favoured related to the teacher solving the problem or showing the correct solution on the board. Even though, in Q9.C, the teacher would correct the solution on the board asking students to generate each step, it is likely that only students who already solved the exercise would participate, and the approach would end up being false instruction, where students who were unable to answer correctly merely copy the correct solution from the board with no guarantee that they understood the mathematical concept.

In Q16 and Q20, which are both related to teacher judgement about the best way of advising students on what to do after feedback has been given, teachers were able to identify the score 4 option.

Therefore, it seems that FD mathematics teachers know what to do when they have to instruct students on what to do next, which was observed when they had to advise students on how to use the quizzes (Q14), and when they recognised the best way of giving advice when students will act upon the feedback by themselves (Q16 and Q20). However, they have a different view when there are opportunities for teacher participation. It seems that they have a need to tell students the correct answer or show them how to correctly solve the problem on the board, moving the conversation to whole class discussion.

This is not to say that a whole group approach should be discarded. In fact, it can be considered a good approach as illustrated in Scenario 1. However, it appears that for a considerable number of FD mathematics teachers this is most often chosen as the best approach, no matter what the situation. In many cases, they did not consider that written feedback had just been given (Q9) and therefore students should be given the opportunity to act upon it (Shute, 2008); that they should value a comment that a particular student has just made (Q7) or yet consider that the misunderstanding that the teacher diagnosed might be the case of just a few students (Q9) and therefore the whole group approach might actually discourage the other students.

Still in terms of the balance between teachers' and students' participation, the items assessing teachers' KLI also provided important information.

There were two questions addressing the learning intentions and one for success criteria. In Q17, which was an example of teachers defining success criteria, the majority of teachers' answers were towards the high end of scores. However, the number of teachers who chose the score 1 option was the same as those who chose the score 3 option. These results show that although most teachers seem to know the importance of students participating in the process of devising success criteria, there is still a considerable number who believe that the best strategy would be to bring pre-defined criteria and explain them to students. This confirms the trend that some teachers are still following a top-down approach, in which students have a passive role of just following what the teacher says.

In Q6, respondents were presented with a situation in which the teacher recognised that there was a problem occurring and had to judge what they could have done to avoid that problem when dealing with the learning intentions. In this example, most teachers were able to identify that they *should have used some strategy to check whether the students understood the learning intentions* written on the board (Hodgen & Wiliam, 2006; Wiliam, 2011), even though a considerable number of teachers believe that *writing more details on the board* would have helped.

On the other hand, in Q12, where teachers were required to judge the best way to check how students understood the learning intentions at the end of a lesson, only 19% of teachers identified the option judged to be the best. Most teachers chose the score 3 option, which was quite similar to the score 4 option in terms of using student-friendly language to talk about the learning intentions, but did not provide teachers with the opportunity to check students understanding. In this situation, Mercer (1995) have shown that the development of mathematical thinking is jeopardised when there is not an attempt to synthesise students' contributions. Teachers seemed to know how to give instructions to students, but need to improve in terms of subsequent actions.

Comparing the results of Q6 and Q12, it looks as if FD mathematics teachers know the importance of checking if students understand the learning intentions and that they recognised when there was a problem with how these were communicated to the students. That is, they were critical when learning intentions were presented to them. However, they did not perform so well when they were expected to refer back to the learning intentions at the end of a unit or lesson.

In addition, teachers recognised the options in which the language was more student-friendly than others. Student-friendly language was reported to be used frequently by teachers in Lysaght and O'Leary's study (2013). Although their study was conducted with primary teachers, and therefore it was expected that teachers would take language into account as they are dealing with younger students, this was the practice that was reported as being the most embedded/established when referring to sharing learning intentions and success criteria.

Lastly, considering the particular characteristics of the items assessing teachers' KAM, the results of these items suggest that teachers are not doing so well in this aspect of FA. There were three items assessing this domain and they all provided important information about what teachers know (or do not know).

First, the results of Q4 showed that the majority of teachers were not able to identify that the activity set by *Mr Fitzgerald* was *not assessing what he wanted to assess*. For some of these teachers the activity was ready to be set (Q4.C) and for others, more exercises were necessary (Q4.D).

Comparing the results of Q4 with those from Q6, it suggests that although the respondents identified the importance of checking whether students understood the learning intentions, they were not able to recognise that an assessment method was not in accordance with the learning intentions. This result resembles Schneider and Meyer's study (2012) in which less than one-third of teachers they sampled showed the skill to properly align students' tasks to the learning goals, which also raised a concern about the purposes of the assessments being carried out.

In this study, teachers' answers to Q4 raised a concern with regards to the knock-on effect this may cause, as poor task design or testing with no clear objectives does not allow teachers to collect the information they need to make informed decisions about how they can adapt their teaching to support learning (Looney, 2011). When comparing these results with how teachers answered the items referring to their KIE, it is possible to question the quality of the evidence that teachers are eliciting through classroom assessments. That is, even though on the one hand it seems that teachers are able to interpret the evidence elicited, this information appears to come from a not so reliable source.

It is essential to remember that in Brazil there are no high stakes external assessments and therefore the results are quite worrisome, as students' futures may be being decided based on unreliable information coming from not-well prepared methods of assessment.

As put by Bennett (2011), what makes something an 'assessment', is not just that evidence is elicited. It requires the careful design of situations and questions so that it is possible to connect the elicited evidence to critical components of domain understanding. The answers to the example included in the MaTFAKi leads to the belief that this connection might not be considered by FD mathematics teachers.

However, as the curriculum in Brazil is structured around topics and not by learning objectives, there may be another explanation for why teachers answered Q4 the way they did. For these teachers, it might be the case that they read the question as *Mr Fitzgerald* creating an activity to assess 'fractions' or 'division by fractions' and not specifically if the students know *how to divide by half.* Therefore, it could somehow be understood, or even expected, that they would consider the activity to be ready to be given to students. The same could be said in terms of the number of exercises – indeed three would not be enough for such a 'large topic'. In any case, this is clearly an area of FA that needs improvement.

Still in terms of their performance on items assessing their KAM, the results of Q11 have also provided important information regarding teachers' views of multiple-choice items and the use of activities to encourage discussion when students are peer-assessing. When asked to judge the best contribution (from fictional teachers) the respondents either favoured *Miss Lee's* opinion – *students focussed their discussions around question 3 because questions 1 and 2 were of multiple-choice* – or did not take sides and chose the option that both teachers presented an incomplete analysis, which was insufficient to explain why students focussed on question 3 only.

In regards to the former, it looks as if the majority of teachers believe that multiple-choice items are not adequate when the activity will be used for students peer-assessment. For them, these items do not encourage discussion and the provision of feedback; instead they believe it is best not to give students multiple-choice items because *the student select an option and that is it*. In relation to the latter, because the MaTFAKi was structured and teachers could not explain their answers, it is not possible to understand what else could have explained the issue with the activity. The only certainty is that for these teachers there is further explanation.

Finally, in Q13, where respondents were presented with a situation in which a teacher was encouraging students to use the results of various quizzes as feedback, the results showed that FD mathematics teachers did not recognise that the purpose was to improve students' learning. Their answers elicited a misconception in relation to a required measure present in the FD official documents: 'the continuous recovery'. In this case, most teachers referred to the continuous recovery as a means to recover students' marks.

Considering that when referring to the continuous recovery the GEA recommends that "if teachers use the evidence from assessment continuously, they will be able to remedy students' problems while learning is happening, which in most cases will make it unnecessary to recover the mark at the end of a period" GDF, SEEDF & SUBEB (2014, p. 24), the results of Q13 show that this message is not reaching teachers as expected.

On the other hand, if for those teachers, 'assessment' is all about 'testing' and the evidence is the resultant mark, it seems logical to expect that they would apply tests more often as a means to "use the evidence from assessment continuously", as the definition states. In this case, it could be argued that teachers are using the continuous recovery in accordance with their view of assessment, and therefore, the problem lies with the definition; not with teachers use of it. Brown (2003, 2008) has showed that teacher beliefs have a huge influence on how they see the purposes of assessment, and Camargo and Ruthven (2014) have shown that tests are usually the most common method of assessment used by Brazilian teachers.

In addition, this may be explained by how assessments are organised in Brazil: as there are no external assessments, summative assessments are the responsibility of teachers. Therefore, classroom assessments become high stakes and a crucial role of Brazilian teachers. As a result, it looks as if teachers are either favouring this function of assessment over the formative one due to its weight, or there is confusion between them, with teachers seeing 'generating a mark' as the only purpose of assessment.

In any of the situations presented above, teachers' answers to the examples assessing their KAM indicated that this is an area that needs further development, which I discuss and suggest some directions for in the next chapter. In the next section, I discuss teachers' answers to the items assessing their KEF.

10.2.2 Knowledge of feedback in particular

Four questions were developed to measure and understand teachers' KEF. The pattern of responses was the same in all four, showing that FD mathematics teachers have divided opinions in relation to the characteristics and the most appropriate type of feedback to be used in a given situation. However, because the questions dealt with different examples of 'providing effective feedback', the characteristics of this division also varied.

When teachers were presented with a situation in which the majority of students were still having difficulties and making the same mistakes (Q2), most teachers identified that the best strategy would be to give oral feedback to the whole group first (Brookhart, 2008) and then *give them another chance to re-do the exercises* (Q2.C). As explained in Chapter 7, this approach was judged to be the best because, in addition to saving time, the teacher could address specific aspects that they identified as important (J. A. Hollander,

2002) and help other students who may still not be very confident about whether or not they know the content (Stiggins, 2009). In this case, a whole-class discussion could help to create learning gains and enhance students' confidence to explain their mathematical ideas (Furtak, Seidel, Iverson, & Briggs, 2012; Kovalainen & Kumpulainen, 2005).

On the other hand, even though the question specifically asked teachers what would be the best type of feedback to give to students in that situation, a noteworthy number of teachers chose the option in which no feedback was given (Q2.B). In this case, it looked as if these teachers believed that being available so students could *ask for the teacher's help when they deem necessary* together with returning the assignment and giving the students the opportunity to re-do them, is a type of feedback and would serve the purpose. This option could be seen as an unspecific strategy, which may leave students uncertain as to how to proceed in the learning process (Kiemer, Gröschner, Pehmer, & Seidel, 2015). Therefore, it is likely that it may not have any positive impact on students' learning.

This could also mean that teachers who judged this approach to be the best may have been thinking that they were encouraging students' participation; allowing students to take charge of their own learning. This would help to develop students' autonomy and lead to a positive impact on learning. However, Labuhn, Zimmerman, and Hasselhorn's study (2010) showed that the self-evaluative judgements of students who received feedback were more accurate than of students who did not receive feedback. The study by Higgins, Hartley, and Skelton (2002) corroborates this idea by arguing that many students read and valued their teachers' comments and adopted a more conscientious approach. They were intrinsically motivated and sought feedback, which helped them to engage with their subject in a deeper way. Likewise, the importance of feedback in helping students to learn how to learn can result in achievement growth and concrete students' participation (Hattie & Jaeger, 1998; Stiggins, 2009).

Similar results were observed in relation to teachers providing written feedback to students (Q8 and Q15). Again, teachers showed a polarised view, this time when identifying how to word the feedback that would be given to students.

For some teachers, their response choice suggested that they understood the ideas that have been supported by the literature (S. Askew, 2000; Brookhart, 2008, 2011; Rakoczy et al., 2013) and identified that effective feedback starts by telling students what they did and then showing what they are missing, helping them to close the gap (Q8.A and Q15.B). The teacher is helping students to understand what they missed (How am I going?), providing a recipe for improvement (Where to next?), and not doing the work for them (Wiliam, 1999). In the examples included in the MaTFAKi, the student would have to figure out that they had to go back to the final answer and analyse what they have to do to get it right. These two options also align to what Santos and Pinto (2009) have shown: the interrogative form (asking a question) in comparison to the affirmative form (telling students what to do) tends to facilitate students' understanding of the feedback and get them involved in the task. Equally, feedback strategies which include allowing students to identify and correct their own mistakes to reach the correct answers, are favourable to long lasting learning (Jorro, 2000; Nunziati, 1990).

On the other hand, as happened with Q2, for these two questions there were a considerable number of teachers that were unable to identify these characteristics and believed that feedback should specifically tell students what they need to do to find the correct answer (Q8.D and Q15.C).

These teachers' choices go against what the literature says in relation to feedback and the teaching and learning of mathematics. Aleven, Ogan, Popescu, Torrey, and Koedinger (2004), for example, have shown that simply receiving feedback that tells the correct answer is least likely to be of value for students, whilst others have found that overly directive feedback is perceived as irrelevant and frustrating (Price, Handley, Millar, & O'Donovan, 2010). Feedback which focusses too much on the mistakes, instead of on how to overcome them, has a negative impact on students seeking more challenges (Hargreaves, 2013; Kay & Knaack, 2009).

It could be argued, however, that with the feedback that these teachers judged to be the best in this case, it is guaranteed that students will manage to find the correct answer, which could be teachers' main purpose. Nevertheless, it is likely that this will happen only because teachers told students what to do (Wiliam, 1999), which does not necessarily mean that they learnt and/or understood why they were following those steps. With this approach, it appears that teachers are taking away the opportunity for students to improve and understand their mistakes through active participation in the search for the correct solution.

It is expected that teachers' approaches should focus on developing a deeper understanding (Donovan & Loch, 2013) rather than merely finding the correct solution or the assimilation of mathematical formulae, and this does not seem to be the case for the teachers who favoured these response options. As noted by Heritage et al. (2009), if the teacher is not able to move learning forward, the value of FA is called into question.

Finally, in a situation of oral feedback (Q19), the results were no different. Some teachers believed that the best way to respond to a student's formulation would be by asking him to explain his thoughts and help him to understand why his initial answer was not in accordance with what the problem was asking (Q19.C). Approaches and feedback that stimulate students' participation, and the explanation of which part of the concepts they know and how they understand them should be favoured in mathematics lessons, and that seems to be the case for teachers who favoured this option.

The idea behind this option is in accordance with Khisty and Chval's study (2002), which found that teachers' focus on mathematical talk and meaning enabled students to develop mathematical reasoning in significant ways when their interaction was less about teacher exposition and more about the perceptions held by students. In White's study (2003), students said that when their teachers actively listened to their ideas and suggestions, it showed that they demonstrated the value they placed on each student's contribution to the thinking of the class.

Likewise, research has shown that verbal teacher-student interactions is one of the major means of constructing meaning (Mercer, 2010; Oliveira, 2010; Webb, 2009). As language use and interaction quality have important implications for students' learning processes and outcomes, active engagement, learning motivation and interest (Sierens, Vansteenkiste, Goossens, Soenens, & Dochy, 2009), what teachers see as constituting effective oral feedback is of vital importance.

On the other side, a number of teachers preferred to tell the student straight away where the mistake was, which contradicts the main principle of FA of using assessment information to improve learning and guide the next steps in instruction (Black & Wiliam, 1998; Kingston & Nash, 2011); and the ideas of meaningful mathematics teaching and learning (Hodgen & Wiliam, 2006).

The findings of this study, therefore, suggest that many FD teachers do not take these principles into consideration and believe the best strategy would be to tell students what would happen if they followed a specific approach (Q19.A) *If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us...* rather than asking him to explain his thoughts (Q19.C). When teachers tell students what to do, instead of providing them with opportunities to understand why they did not get the correct answer, they are taking away students' opportunity to learn. They will get the correct answer, but this will be because they were just told what to do.

Even though with this approach it is possible to convey that the teacher values classroom discussions, as they spoke to the student and encouraged him to *find another solution*, Black and Wiliam (1998) found that not all classroom discussion and questioning results in improved learning. When the teacher focusses on questioning to lead the student to a particular (expected) answer, it inhibits their learning as it takes away the opportunity of coming up with their own answers that were unanticipated (Davis & McGowen, 2007).

It is important to give students the opportunity to express their mathematical ideas, as Anthony and Walshaw (2002) found that many students do not know how to explain them. Indeed, in their study, several students were afraid with the idea that they would have to share their thinking with others. Equally, Bicknell (1999) showed that many of the Year 11 students who participated in her study on mathematics assessments were uncertain about what was expected when asked to explain their ideas.

The cognitive advantages to students through their participation in mathematical discussion have been reported by many authors (e.g. Boaler, Wiliam, & Brown, 2000; Nathan & Knuth, 2003; Woodward & Irwin, 2005). As said by Walshaw and Anthony (2008, p. 526),

> By expressing their ideas, students are able to make their mathematical reasoning visible and open for reflection. Not only does the expression of student ideas provide a resource for teachers, informing them about what students already know and what they need to learn, the ideas also become a resource for students themselves – challenging, stimulating, and extending their own thinking.

The importance of talk in mathematics teaching and learning should not be ignored, but encouraged (Hodgen & Marshall, 2005; Webb, 2009). Practices that create opportunities for students to explain their thinking and to engage fully in dialogue have been reported in research undertaken by Steinberg, Empson, and Carpenter (2004).

In the same direction, guidelines have been provided and they defend, among other things, that effective teaching involves observing students and listening carefully to their ideas and explanations (NCTM, 2000; GDF, SEEDF & SUBEB, 2014). Yet – and the response option chosen by teachers could illustrate this – classroom interactions lack cognitive spaces as they are often dominated by close-formatted classroom talk, especially in regards to STEM-subjects, with negative impact on students' long-term development (Jurik, Gröschner, & Seidel, 2013; Walshaw & Anthony, 2008; Wells & Arauz, 2006).

This lack of cognitive space is not an issue only for FD teachers. It was observed with teachers in the other states of Brazil (although in a small proportion) and in other countries. The study by Woodward and Irwin (2005) focussed on two teachers' verbal interactions with their students. The results showed that cognitive space was limited by too much feedback, the lack of pause times for thinking, and students being occasionally "talked over" (p. 802). Specifically, the authors noted that students did not have the opportunity to learn and speak the language of mathematicians. They concluded that by not providing students with opportunities to engage in mathematical discourse, the teacher prevented students from expressing what they were learning, which corroborates Anthony and Walshaw' (2002) findings discussed above.

This constraint of cognitive space was also observed in the study by Khisty and Chval (2002). Although the teacher contextualised mathematics in a story that had relevance to students or developed a role-playing activity to assist with conceptual understanding, the teacher nevertheless failed to provide students with opportunities to "talk the talk" (Hodgen & Wiliam, 2006, p. 4) associated with mathematical knowledge.

Lastly, in all four questions, if the answers are analysed in two blocks (scores 1+2 and 3+4 together) there is almost a 50-50 division. This could be interpreted that the results are not so alarming when compared to teachers' performance in other aspects of FA. However, considering feedback as the heart of FA, and its importance for the development of students' learning and self-regulation, and for teachers' practices – this is an aspect in which the results might have been hoped to be better.

Even though research has shown that providing effective feedback is a hard skill for teachers to acquire/develop (Brookhart, 2011; Donovan & Loch, 2013; Nicol & Macfarlane-Dick, 2006; Wiliam, 2011) and to put into practice in their lessons (Black et al., 2003), this cannot be used as an explanation for teachers' performance in this study. Teachers were required only to identify the best feedback among four options, and not actually to provide it to students in a real situation. This suggests that indeed there is a gap in teachers' knowledge in this aspect and therefore this is something FD mathematics teachers should be focussing on developing and improving.

10.3 Answering the second set of research questions

Are there any systematic patterns of variation in teachers' knowledge:

1. relating to different aspects of FA?
2. according to background characteristics of the teacher (e.g. years of teaching experience or level of education)?

10.3.1 Variation of teachers' knowledge relating to different aspects of formative assessment

Table 9.3 suggests that there were three aspects of FA that presented a systematic pattern of variation in teachers' knowledge relating to different aspects of FA.

The first two aspects are in relation to all questions assessing teachers' KIE and KHS, which were considered as part of teachers' strengths because their answers were concentrated in the high end of scores and therefore showed that in the examples included in the MaTFAKi teachers did relatively well. The nuances of each of these items and the immediate impact of them in mathematics teaching and learning were discussed in Section 10.2, because they were answers to the first research question. These two DoK have proven to be strengths not only of FD teachers, but in Brazil as a whole (Table 9.4).

The third aspect is in relation to teachers' responses to the items assessing their KEF. As with the KIE and KHS, the specific knowledge assessed by each item of the KEF was discussed in Section 10.2.2. They were included in the 'divided performance' category because the participants could be split into two groups: *1*) those who identified the answers considered to be the best and whose KEF could therefore be considered a strength and *2*) those who did not identify the best answers and therefore their KEF could therefore be considered a weakness. As the same pattern was not observed with the *BR* teachers (Table 9.4), the results convey that this is a systematic variation of a particular aspect of FA knowledge of FD mathematics teachers, but not of those in the sample of the other states of the country.

It is important to emphasise that items were grouped based on their patterns of responses and not by the results of the factor analysis. However, even though the results of the factor analysis did not group these items into factors, the results discussed here and in the previous sections still provide important information, especially considering teachers' KEF, KAM and KLI, which showed that in the situations represented with the items in the MaTFAKi, there are a great number of teachers that are not doing so well in these aspects of FA. Considering that they only had to choose among four options and were still unable to identify the best action to take, it leads to the belief that the situation could be even worse if their actual practices were investigated.

10.3.2 Variation of teachers' knowledge according to background characteristics of the teacher

Differently from above, the findings presented in Section 9.5 have shown that only a few unrelated questions (Q11, Q16, Q17 and Q20) produced statistically significant results when comparing the percentage of teachers choosing each response option in terms of teachers' years of experience, their level of education or whether they have or not the teaching qualification in mathematics.

The generally null results across items, then, could be interpreted as showing: *1*) that there is no substantial difference in knowledge of FA of mathematics teachers in the FD according to these factors, and so that such knowledge is not being influenced by teacher education or classroom experience; or *2*) that the MaTFAKi was not able to capture any such differences and so provide this information.

In any case, further investigation is necessary to better understand what shapes teachers' knowledge of FA to be able to make any claims with certainty.

10.4 Discussing the MaTFAKi

As explained in Chapter 4, I ended up having an additional goal of analysing:

> To what extent was it possible to produce an instrument which would reflect, in a practical way, what has been said in the literature of FA?

Therefore, in this section I discuss the MaTFAKi as an instrument to assess mathematics teachers' knowledge of FA. This involves the process of developing the MaTFAKi and what FA means in mathematics classrooms in the FD, its validity and reliability, and which further measures can be taken to improve the MaTFAKi.

10.4.1 The process of developing the MaTFAKi

As explained in Chapters 5 and 6, the MaTFAKi was carefully designed, incorporating various stages of piloting and revision. Some versions are provided in the appendices, but many more iterations were necessary to come up with what was considered to be its final version. This included not only considering the content of each scenario and answer options, but also issues of language, position of options, consistency of terminology, and length of the MaTFAKi, among others.

Being the first instrument of its kind, and due to the complexity of FA, turning the theoretical recommendations into classroom scenarios and questions with plausible response options was not an easy task. The main purpose of using a scenario-based instrument was to truly represent classroom situations that were not only relevant to FD mathematics teachers but would also take into account the recommendations of the research literature also encouraged by the institutional literature.

Finding situations that could be considered FA examples, or even coming up with them from my own experience as a teacher, was not so hard. The trickiest part was to turn them into situations that were relevant to the Brazilian context and were 'closed' enough that it would be possible to affirm that there was a 'best' answer.

In this direction, I tried to include different situations of mathematics classrooms taking into consideration various aspects of FA such as:

- different topics from the curriculum at different stages (equations and inequalities, fractions and percentages, circles, surds, etc.)
- the interpretation of students' learning based on written exercises and/or oral comments
- the teacher giving feedback orally and in the written form – to the whole group or individually
- feedback between teachers, between students and to parents
- examples of strategies that teachers can make use of to implement FA in their lessons, including peer-assessment or summative tests with formative purpose, among other things.

There were aspects of FA that, to my knowledge, were not as likely to happen in Brazilian classrooms, but were included due to their importance to FA. This was the case, for example, with the use of learning intentions and success criteria. As the Brazilian national curriculum only specifies the content to be learnt and it is not organised in terms of learning objectives or intentions, teachers do not prepare their lessons based on the learning intentions and tend to follow the sequence of the textbooks (Dante, 2008).

Being relevant to Brazilian teachers also meant, for example, including mathematical tasks that are not in accordance with the latest recommendations for effective mathematics teaching and learning (Bishop, Clements, Keitel, Kilpatrick, & Laborde, 2013; English & Kirshner, 2015), but that would resemble the tasks teachers are used to setting for their students. On the other hand, it meant including aspects that were relevant only to the FD context, which was the case of the 'continuous recovery'.

In this sense, being Brazilian and having experience of the FD system supported the development of the MaTFAKi in such a way that it would provide valuable information on teachers' misconceptions and the distance between what the official documents preach and what teachers understand and probably put into practice.

However, as I could be wrong or outdated since I have been away from Brazil for some time, in every piloting phase I made sure to ask participants that I was not including anything that they would not do in their lesson (see Section 6.2) and if I was leaving out anything important.

Hence, developing the scenarios and the questions embedded in each scenario, was not just a matter of following the literature, but of actually finding the right balance between what has been said in the literature, what was likely to happen in Brazilian classrooms, the language to link them both in a way that teachers would recognise, and more importantly, if it was assessing the DoK that it was supposed to assess.

However, even when all these goals were met, some other elements had to be taken into account, meaning that scenarios and questions to be dropped or redeveloped. One of these elements was the length of the MaTFAKi and how long teachers were taking to answer it. Being careful about the length meant not only reducing the number of questions but also how wordy the scenarios and answer options were. Adding to this, as the target population was Brazilian teachers, producing the Portuguese version was facilitated by the fact that it is my native language, but made it harder to keep it succinct as written Portuguese can be very wordy and different from the English language. Some translation had to be explained in brackets due to unfamiliarity to teachers and to prevent different interpretations, as it was the case with 'peer-assessment'.

Dropping questions was not just due to being careful about the length of the MaTFAKi. As specified in Section 6.3.2, four questions were dropped after the field test. Therefore, even though these questions were considered to be relevant to teachers and were judged to be aligned with the designated DoK, they were dropped as not being as reliable as the others. Consequently, some scenarios ended up having more questions than others and some DoK were assessed by more questions than others, which was not the initial plan. Information about teachers' KHS, for example, was provided by only two questions.

Equally, even though the field test had identified some options as not being as popular as others, I had to decide to move forward and send the MaTFAKi to FD teachers because there was no more time available for further piloting. Re-writing one option or one question would mean going over most of the validation phases and field test again which was impracticable in the time frame available.

Yet, considering the practical elements discussed above and the static nature of a questionnaire, it was not possible to include some aspects of FA. One example is the inter-personal nature of FA. That is, even though teacher-teacher, teacher-student(s) and student-student exchanges were exemplified in the MaTFAKi, it could not encompass real human interaction – there were not 'in the moment' exchanges and therefore could have produced different results if analysed in real-life situations.

However, this was a decision that was deliberately taken due to the results of previous research (Albuquerque, 2012; Camargo & Ruthven, 2014; Camargo, 2015), my attempted research explained in Chapter 1 and my experience of Brazilian mathematics lessons.

In conclusion, the MaTFAKi was carefully designed to fulfil the purposes of the research and the discussion above shows that it is safe to say that it was accomplished even though some improvements can be made. In the next section, I discuss how the development of the MaTFAKi and the results of various qualitative and quantitative analysis show the validity and reliability of the MaTFAKi and therefore corroborates the points made in this section.

10.4.2 Validity and reliability evidence for the MaTFAKi

Just as it is claimed that teachers should be able to draw accurate inferences from test results, it is important that the inferences drawn about teachers' abilities are equally accurate. To make these inferences, the research presented here comprised reliability analysis and a thorough validity investigation of the MaTFAKi content and scores coming from MaTFAKi responses. Overall, tentative support for the MaTFAKi was found.

Teachers opinions and understanding of the items were taken into account more than once and the results of numerous face-to-face review iterations have shown that items were correctly interpreted by teachers and accurately reflect understanding of the targeted concepts/actions. When that was not the case, further revisions were conducted to reflect their views. Even though it could be argued that only a few teachers had the chance to 'validate' the MaTFAKi, the sample of teachers was carefully chosen to include teachers teaching in different schools, from various regions within the FD, and in different stages of their career. Measures were taken to guarantee that their views reflected those of a larger pool of teachers and that the MaTFAKi was written in a familiar vernacular, with scenarios and questions likely to happen in their actual classrooms.

The same can be said about the content of FA, the ranking order of the options and the explanation of each item. Various revisions with the participation of early career and senior researchers were conducted to consider that FA was well-represented not only by the six DoK chosen as the framework for the MaTFAKi development, but also that each question was assessing what it was intended to assess. The researchers were invited due to their knowledge of assessment, mathematics education and the FD educational system. The examples explored in Chapters 5 and 6 show the extent of discussions that happened with the researchers.

Even though I acknowledged in the previous section that there were some aspects of FA that were left out – most of it for practical reasons – the MaTFAKi can still be considered a valid instrument to assess the knowledge of FA of FD mathematics teachers.

In terms of reliability, the high results of each of the 20 questions as yielded in the test-retest analysis (see Section 6.3.2.2) permit the affirmation that the MaTFAKi produced consistent results (Multon, 2010; Rust & Golombok, 2009).

It is possible to say that most of the MaTFAKi items function well, while only a few need revision. This was observed in terms of some response options in particular, but also in questions (as a whole). Q1.C, Q7.A and Q13.C are examples of the former. These options were chosen by less than 2% of participants, which shows that they were not plausible options. Analysing the content of these options further, some tentative explanations can be made.

In Q1.C, as it is clearly stated that students *do not know* how to solve inequalities, it might have put teachers off due to how it was written. Perhaps if instead of writing *do not know*, *might not know* could have attracted more teachers. Q7.A may have had the same issues as in Q1.C, in that it does not represent the teachers' view.

On the other hand, with Q13.C, it seems the issue was due to its similarity to Q13.D, and therefore teachers ended up favouring Q13.D which was a 'better' version of Q13.C. Therefore, I believe it is possible to say that in Q13 the lack of plausibility of the option was a matter of not noticing the similarity with another option; whereas in Q7 and Q1, the options may not reflect the teachers' views.

Yet, I believe that these options still provide valuable information about teachers' knowledge. In Q1, participants have shown that they do not consider only the final answer to judge whether a student has or has not learnt the content. In Q7 it shows that teachers recognised that, in the situation presented, students needed some help or guidance and that just asking them to *discuss some more* would not be enough. Finally, in Q13 it seems that those teachers who did not favour the option mentioning 'continuous recovery', identified that in Q13.D a better mark would be the result of *learning*, whereas in Q13.C it would be due to *practising*.

Q11 is an example of an item that needs revision (as a whole). The results of this question were very different from the others in the MaTFAKi and the opposite of the results in the *BR*. Being the only item where most teachers' answers were concentrated in scores 1 and 2, it could be considered that some kind of bias was introduced and may reflect the opinion of only the sample who answered the MaTFAKi in the FD.

Even though the results presented in Table 6.14 guarantee the reliability of this question, the retest was not conducted with the sample in the FD and therefore it might be the case that this question encompasses a content that is either more relevant for teachers in the other states of Brazil or shows that FD teachers may have a very opinionated view in terms of setting multiple-choice questions.

Finally, claims regarding the general level of a teachers' FA knowledge, or the general level of knowledge of a population of mathematics teachers, can be supported with evidence from the various analyses that were conducted and discussed in the previous sections.

Claims may also be made in regards to systematic patterns of teachers' responses in the FD, as discussed in Section 10.3 and when compared to the results of the other states in Brazil, but finer distinctions of the knowledge levels based on the DoK may not be supported, given the results of the factor analysis.

In the next section, based on the discussion of the two last sections, I explore some ideas on how to improve the MaTFAKi.

10.4.3 Improving the MaTFAKi

Assessing knowledge of FA through a structured questionnaire has its limitations. However, I view these as a way to improve the MaTFAKi which I intend to do in the near future in collaboration with others.

First, as it is likely that teachers did not choose some options because they did not reflect their views, an immediate improvement would be to replace the response options that had a low response rate to make them more plausible and attract more teachers.

Second, as discussed in Section 10.4.1, some aspects were left out and therefore there is a need to develop more questions so those aspects can be assessed using the MaTFAKi. This would be facilitated if the development was done in conjunction with a team of researchers, for example, through further collaboration with the group in Brazil that participated on the validation phases.

Third, considering that only teachers who participated in the validation phases had the chance to justify their answers, another way of improving the MaTFAKi and consequently the analysis of its results would be to include opportunities for teachers to explain or justify their choices after each question. This would allow for an extended assessment of teachers' knowledge of FA, with a better understanding of their views.

However, in the last two suggestions the length of the MaTFAKi and the amount of time required for teachers to answer would increase considerably. This would also reduce the likelihood of voluntary participation. Therefore, a fourth suggestion would be the application of Item Response Theory (IRT) instead of Classical Test Theory (CTT) as the former provides more in-depth analysis of the items and allows the development of efficient questionnaires by reducing the number of items to be included (Reeve & Mâsse, 2004). The online feature of the MaTFAKi facilitates this measure to be taken.

10.5 Summary

In this chapter I have discussed the main findings of this study, dividing them into two parts.

In the first part, I discussed mathematics teachers' knowledge of FA as measured by the MaTFAKi. As I hypothesised, it was possible to observe that in some aspects FD mathematics teachers already understood the ideas that have been supported by the literature but in some cases the gap still exists.

It was also possible to observe some systematic patterns of variation in teachers' knowledge in relation to different aspects of FA, but not according to their background characteristics.

The results discussed in the first part shows that there are several aspects that need to be addressed when referring to mathematics teachers' knowledge of FA in the FD due to the impact they have on the teaching and learning of mathematics, as well as on the impact of this on their students' lives. The discussion in the sessions above has shown how a failure in one of these aspects can reduce the effectiveness of FA.

In the second part, the MaTFAKi as an instrument to assess FD mathematics teachers' knowledge of FA was discussed and although some improvements could be made, the MaTFAKi is a valid and reliable instrument for its intended purpose and use.

In the next chapter, I summarise the main findings discussed here and conclude this dissertation, presenting the implications for policy, practice, and future strands of research.

CHAPTER 11

CONCLUSION

Throughout the dissertation, I argued that my purpose with this study was to better understand mathematics teachers' knowledge of FA in the FD of Brazil so that I could feed back to them, and those working with them, on how to improve. According to Hattie and Timperley (2007) effective feedback provides answers to three main questions: 'Where am I going?' 'How am I going?' 'Where to next?'

The information to answer 'Where am I going?' was provided in Chapters 3 to 7, where I outlined some essential aspects of FA that teachers should know and explained how these aspects were assessed with the MaTFAKi. 'How am I going?' was answered by the information from Chapters 9 and 10, where I presented and discussed teachers' knowledge of FA based on their answers to the MaTFAKi.

In this chapter, my purpose is to provide information to answer the final question 'Where to next?' to fulfil the main goal of this study.

Section 11.1 summarises the main findings discussed in the previous chapter. Section 11.2 suggests implications for policy and practice, based on the main findings. Section 11.3 outlines the limitations of the study, in addition to the ones presented in the previous chapter. Section 11.4 explores possible directions for future research. Section 11.5 concludes this dissertation with final remarks on this study.

11.1 Summary of main findings

Reflecting on the results from the data generated by the MaTFAKi and presented in the previous chapters, I summarise the main findings below:

- The MaTFAKi, as designed and developed, is fit for purpose; findings from the validation phases and statistical analyses confirm that the questions are suitable to assess the knowledge of essential aspects of FA of mathematics teachers in the FD of Brazil.

- The MaTFAKi provided a snapshot of the current state of mathematics teachers' knowledge of FA in the Brazilian FD. In doing so, it provided information that was not available and therefore enriches understanding of mathematics teachers' FA ability.

- This snapshot has suggested that FD teachers' general level of FA knowledge is lower than that of teachers in the other states of Brazil.

- This snapshot has also shown that there are aspects that teachers are doing well and others that they need to improve. Participants showed a good performance in terms of interpreting evidence of their students' learning and helping their students to use assessment information. On the other hand, teachers had a relatively low performance in terms of choosing/developing assessment methods (classroom activities, discussions) to elicit evidence of students' learning.

- In terms of 'providing effective feedback', FD mathematics teachers have a divided view. Some teachers recognised that effective feedback provides students with information to close the gap in their learning without doing the work for them, but an almost equal number believed that effective feedback tells students the correct answer or exactly what they need to do to get to the correct answer.

- The results of some items provided valuable information about participants' misconceptions and some inconsistencies between what the official documents state and how teachers understand it. They also showed that FD mathematics teachers tend to favour some strategies no matter the situation (e.g. showing the correct solution on the board and whole-class discussions) and still do not recognise others (e.g. peer-assessment).

- Finally, there were no systematic patterns of variation in teachers' knowledge according to background characteristics of teachers. Only a few separate questions have shown statistically significant differences when comparing the answers of teachers with different years of experience, level of education, or with or without the teaching qualification in mathematics.

These findings have direct implications for policy and practice, which I explore in the next section.

11.2 Implications for policy and practice

This section presents some implications of this study. The aim is to translate the data into usable information for different purposes and users (AFT, NCME, & NEA, 1990; Brookhart, 2011; Dudley & Swaffield, 2008; Laveault & Allal, 2016; Stiggins, 2011).

First of all, the design of the MaTFAKi can be considered the main contribution of this study to the broad research on FA. The thorough explanation provided in Chapters 5, 6 and 7 shows the conceptualisation of FA in this research and its operationalisation through the development of a structured questionnaire with scenarios of mathematics lessons. These chapters also provided reference to validity literature, explained my phased approach to testing validity and a detailed explanation of the questions, their scoring and the explanations of the scores, showing that the MaTFAKi is a valid and reliable instrument to be used of further research on FA.

Developing the scenarios and the questions embedded in each scenario, was not just a matter of following the literature, but of actually finding the right balance between what has been said in the literature, what was likely to happen in Brazilian classrooms, the language to link them both in a way that teachers would recognise, and more importantly, if it was assessing the DoK that it was supposed to assess. However, by opting for a structured questionnaire as the method of data collection, only a scenario-based type would be able to capture nuances of teachers' knowledge of FA.

With these characteristics, the MaTFAKi can also be considered a great tool to be used for teacher education, even though it was initially designed for research purposes. The scenarios and embedded question in the MaTFAKi can be used in the future with pre- and in-service teachers as a starting point for problematising classroom situations and provide a rich environment for discussions about the several aspects of FA and the best action to take.

Although the scenarios and questions in the MaTFAKi are still hypothetical and limited in terms of variety and number, I am confident that they are grounded in FA and mathematics learning and teaching issues that previous research and experience have highlighted as fundamental and/or likely to occur in actual mathematics classrooms. The phased development of the MaTFAKi provided several opportunities for academics and teachers to reflect on and discuss about knowledge of FA, subject matter, students and learning and the (often teacher-centred) teaching of mathematics.

Equally, with the MaTFAKi, it was possible to assess information that quite likely would not be observed (or mentioned) by teachers in their lessons or considered as FA practices. The reasons and advantages for using a questionnaire was explained in Section 4.2.1, one of which is that it was possible to choose the aspects that were considered to be essential FA knowledge and could provide fundamental information about teachers in Brazil that is not available anywhere else.

In doing so, it is safe to say that the MaTFAKi indeed provides this information and therefore fulfils the purpose of the research by feeding back to teachers and those working with them. I also believe that, in contrast to posing questions at a theoretical, decontextualised level, by using classroom scenarios, the MaTFAKi successfully generated significant access to teachers' views of and actual FA practices.

It is important to highlight, however, that when structured questionnaires are used, the answers are pre-determined and respondents are requested to position themselves, creating a situation that differs from others in which the teacher can talk freely about how they work with their students, what kind of activities they set, etc. As a result, it was not possible to make specific conclusions, centred on each participant, because "there is not an opportunity to work with the particulars, with details, with nuances that in fact characterise each participant and his/her opinions" (Sztajn, 2000, p. 226). The use of a questionnaire produced a panoramic snapshot of a set time and space, on my chosen focus, or on what I included as the items and their four answer options. In this sense, the results provided an understanding of the group of teachers, and not of each teacher separately; and also of a group that positioned themselves based on the answer options that they faced within very specific contexts.

These collective findings show that it is of utmost importance to stop considering that "teachers know much of what they need to know already" (Wiliam, 2007, p. 201); in the MaTFAKi, FD mathematics teachers were presented with only four options to choose from, and some teachers were still unable to identify the best strategy to follow in the situation presented. With the belief that teachers' actions are always towards doing the best possible job for their students, I take the risk to conclude that those aspects were areas in which teachers lack the knowledge.

While this 'lack of knowledge' could be a consequence of initial teacher training, and considering that much of the learning on FA can barely be achieved during the teacher education years and will primarily need to be supported and reinforced over many years (Laveault, 2016), I share Wiliam's (2013) opinion that it is necessary to take care of current teachers – the ones that are already teaching. Therefore, this suggests that there are some immediate implications for policy in the CPD offerings in Brazil.

Even though there is not a general agreement about the theoretical basis of FA specifically (and teaching and learning of mathematics in general), having the opportunity to reflect on and discuss these aspects with colleagues and specialists is important because it provides coherence, helping to ensure that separate elements of effective practice make sense and work together (Laveault & Allal, 2016). Therefore, it should form part of a larger activity network that allows students to develop habits of mind to engage with mathematics productively and make use of appropriate mathematical tools to support their understanding (Walshaw & Anthony, 2008). Even though teachers and teacher education students may have little or no patience with theory, theory is especially important when attempting to change teaching practices, as it can be a guide of what to do when past experince does not shed a light. The knowledge base for effective FA practice is essential (Laveault, 2016; Maclellan, 2004; Xu & Brown, 2016). FA in mathematics lessons will only demonstrate positive effect when there is a strong knowledge base cohesion between all the various elements and interrelated aspects of a teacher's work.

Therefore, rather than imagining that learning about FA would need to be a new, entirely separate initiative, policy-makers in Brazil might consider building FA ideas and processes into the existing CPD offerings. In this way, both the theory of the reforms and the specific teaching strategies would be more coherently tied together for those teachers attempting to try out these reforms in their lessons (Laveault, 2016). At the same time, it would not present FA as a separate 'bolt on' activity.

With the new national curriculum[1] about to be launched, now is the time to develop a strategy that encompasses FA and the shared role and responsibility for learning: a

[1] http://basenacionalcomum.mec.gov.br/images/BNCC_publicacao.pdf

strategy which brings together a deeper understanding of principles and practices that develop mathematical understanding rather than the simple recall of knowledge. Given that for many students mathematics can be a challenge or a synonym of continuing failure, policy that encourages the development of FA in mathematics classroom in Brazil is necessary as it has already proved to bring positive outcomes on students' learning and "is more cost-effective than any other strategy" (Wiliam, 2013, p. 184).

Considering the importance of having assessment methods linked with the learning intentions and success criteria that were shared with students, the launch of the new national curriculum might be the opportunity to bring these elements together as it is now based on learning objectives and will probably make more sense to teachers, consequently helping to strengthen their knowledge in this aspect which the findings have shown to be an area of relative weakness.

On the other hand, it is important to re-consider the top-down, one-off workshop model currently in place in Brazil and how the messages are being delivered through the official documents. Policy-makers need to be aware that teachers will not change if they are told what to do. They need to become active agents and the owners of the change.

In my own view, informed by my teaching experience and from working with Brazilian mathematics teachers, any effort towards implementing FA in their lessons should start by providing teachers with opportunities to learn about and discuss FA (Engelsen & Smith, 2014) and then move to the implementation of some practices, with their own needs as starting point (Black et al., 2003). Previous research (Camargo, 2012) showed that Brazilian mathematics teachers have particular interest in learning how to make use of different methods to assess students' learning and how to use self- and peer-assessment to include the students in their own assessment.

Even though I reiterate that no aspect is more important than any other, there is the need to start somewhere. In this study, the results suggest that some initial aspects of FA need to be taken into consideration before FD mathematics teachers are able to understand feedback as the heart of FA. Figure 11.1 (derived from Figure 5.1) highlights which aspects I believe should be tackled first.

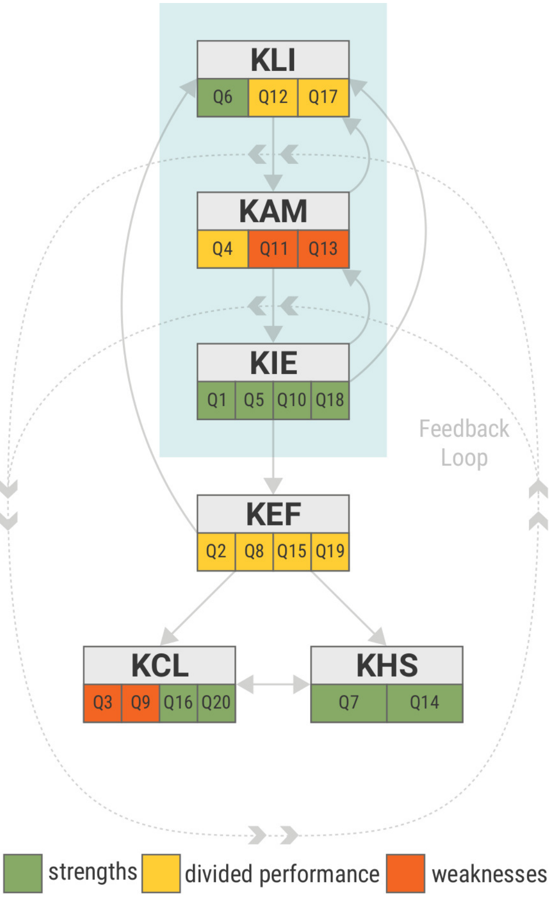


Figure 11.1: Illustration of findings.

The results suggested that teachers did not do so well in the items assessing their KAM. Therefore, it seems logical that any policy developed to tackle teachers' knowledge or use of FA should start by addressing their assessment methods.

However, I believe that focussing only on one aspect may give the false impression that improving an isolated aspect would be enough. Equally, attempting to improve all aspects at once could overwhelm teachers with too many concurrent changes and knowledge to be acquired. Therefore, I suggest not only focussing on their assessment methods, but on the three DoK grouped in the blue shaded area in Figure 11.1 for several reasons.

First, FA is a process that encompasses more than one DoK in which the main idea is to seek and interpret evidence for use by students and their teachers to decide where students are in their learning, where they need to go and how best to get there (ARG, 2002). Since FD mathematics teachers may be struggling to 'seek' it properly, poor performance in KAM potentially jeopardises the rest of the process. For example, interpreting evidence collected through a not well-developed assessment method will end up not being a valid interpretation of students' learning. Equally, giving feedback based on that same information interpreted from a non-reliable source, will result in ineffective feedback.

In addition, with the fundamental role that teacher assessment plays in the Brazilian context, collecting non-reliable evidence from students' learning impacts not only with what teachers are doing in their lessons (formatively), but also in terms of (summatively) deciding their students' future.

Second, since the MaTFAKi identified some aspects in which teachers are doing relatively well, any strategy to support teachers should build on the knowledge they already have, and show how their existing knowledge can be put to better use if other aspects are also strengthened.

Third, the learning intentions are the starting point from where teachers should plan their lessons and should be linked to their assessment activities. If teachers' KLI is not well-developed, it is highly unlikely that the rest of the process will follow in a thoughtful and reflective way. As FD mathematics teachers already demonstrated understanding of some aspects of KLI, supporting them to develop this understanding even further will help strengthen the other aspects as well.

As mentioned above, with the launch of the new national curriculum, which is now structured around learning objectives, it seems to be a good time to think about a strategy that would help teachers implement the new curriculum with FA embedded in their practice. The change from a content-based curriculum to one structured with learning objectives, shows that the government recognises their importance.

Fourth, in the MaTFAKi teachers showed a strong KIE, therefore it seems a good idea to support them to focus on how to choose or develop assessment methods that will produce quality information. Consequently, they can use their strong KIE to make well-informed decisions about any next steps in teaching and learning.

Even so, supporting teachers to develop their knowledge of FA is not enough. Teachers need to see the value of FA as a means of helping students learn (Black et al., 2003) and develop as autonomous learners (O'Shea, 2015). They need to observe others using the principles in practice, and therefore, time and resources must also be spent in developing teacher networks that focus on building reflection on teaching and learning within their own context. For some teachers this will mean a shift to understanding and accepting that FA cannot follow a teacher-centred approach but students taking ownership of their learning.

This shift also means that any policy should enable, encourage, allow and support teachers to change their practices, and should involve principals, students, parents and the whole community, ensuring that those within the community are given opportunities to talk about, support, and nurture each other's development. Educational change requires change beyond surface structures or procedures, focused on altering the knowledge, skills, attitudes and beliefs of all involved agents (e.g. Coburn, 2003; Engelsen & Smith, 2014).

However, as said by O'Shea (2015), even when knowledge is acquired, teachers and the whole education community need to understand that any changes in classroom practices takes time and ground rules are needed so the changes can be transferred to how students perceive assessment and the learning of mathematics. These rules should be negotiated, but need consistent reinforcement over the year. Teachers will require persistence and patience instead of quicker solutions. Telling students about their learning is certainly quicker than enabling students to analyse their own misconceptions, which will impact on how they provide feedback. Equally, if FD mathematics teachers do not truly believe that students can be the owners of their learning, it is highly unlikely that students will believe it too.

When these changes are not supported by external policies, or quality CPD is not available, teachers should look for ways to improve and seek support within their own school or community, as collaborative learning appears to be an essential condition for sustained and durable learning of FA (Laveault, 2016). Especially in the FD, where teachers have dedicated non-teaching time, they can look for ways to develop by sharing their practices, observing each other's lessons, planning lessons together, etc.

In summary, the main implications of this study suggest that the need for FA education for mathematics teachers in Brazil is urgent. Based on current research and the results of this study, for Brazilian mathematics teachers FA education needs to encompass both theoretical knowledge of assessment and more consciousness-arousing components that will prompt teachers to re-examine their conceptions of mathematics teaching and learning.

11.3 Limitations

In the previous section I have discussed some implications of the results of this study. In this section, I acknowledge some limitations. The limitations here do not include the limitations of the MaTFAKi as they were already discussed in Section 10.4.3.

First, due to the lack of research in the Brazilian context, the ‘theory’ in which I based the explanations and rationale for developing the MaTFAKi and the response options, as well as the scenarios and examples, were all based on my own experience and on research outside Brazil. Therefore, although the six DoK chosen to comprise the framework representing essential aspects of FA (as well as the questions generated from them) were validated by particular teachers and researchers in Brazil, it might be the case that some examples included do not reflect the wider experience of Brazilian teachers and this, to some extent, might have influenced teachers’ responses.

Second, only teachers in the validation phases had the opportunity to justify their answers. It may be the case that teachers who participated in the final survey had a different and reasonable explanation for choosing one option over the other but, due to the structure of the MaTFAKi and the distant approach of the research, did not have the opportunity to explain their reasons.

Still, it is important to emphasise the complexity of FA, in that it involves a lot of elements and is contingent on a range of cultural scripts and imperatives. Each classroom context brings its own characteristics to the wider network of FA and pedagogical activity systems and it may be the case that the MaTFAKi was not able to capture all these nuances. Therefore, the use of different methods, such as interviews might have been beneficial for the understanding of some teachers’ responses.

Finally, I acknowledge that this was an ambitious project for the limited financial and human resource and time available for a PhD project. The opportunity to develop more questions and further piloting stages would benefit the validity of results. A larger sample size would also facilitate the factor analysis and the comparison of responses according to the teachers’ background information.

In terms of sample size, the web-application Qualtrics® logged that 577 teachers started answering the MaTFAKi, which means that the method of delivery worked. However, considering that only 169 of them finished answering, something made them give up. Unfortunately, due to the anonymity of answers, I can only hypothesise: the respondents did not belong to the target population and when they realised, they stopped answering; they did not have enough time to finish in the first go and ended up not going back to it afterwards, either because they forgot or did not want to; or they felt that the questions were not relevant and did not see the purpose of participating in the research.

Yet overall, I have attempted an innovative approach in this research and as such, I acknowledge that with the development of a new instrument, some results are tentative. My motivation for this was the lack of research about FA in Brazil and, as a consequence, the lack of information on FA in mathematics education. With this in mind, I believe this study has revealed some interesting aspects that could be turned into future research, which I explore in the next section.

11.4 Suggestions for future research

Reflecting on future research brings several points to mind. This includes not only how to improve and broaden my own research, but more importantly how this research can influence research in both the Brazilian and international context.

First, taking into account the mentioned MaTFAKi improvements; the above limitations; and the fact that the test-retest reliability was high, but the internal consistency and the factor analysis did not yield satisfactory results, there is a need for subsequent (replication) studies to test the consistency and generalisability of these initial findings. The MaTFAKi can be used in other contexts, which will mean that the results can be compared and will allow specific characteristics of FA in mathematics lessons in Brazil to be better understood.

Second, the natural next step is to conduct observations to compare teachers' practices with the results from this study. While measuring aspects of teachers' FA knowledge should be the starting point, as it provides an estimate of their knowledge base (Xu & Brown, 2016), transfer to classroom practice is not guaranteed. Future research on the relationship of teacher knowledge and actual teaching behaviours is needed, especially in the Brazilian context. Further work is needed to explore how teachers' knowledge is related to job performance, teaching effectiveness and how this knowledge is influenced by school-based and educational policy.

Third, considering the active role of students in FA, further research can be conducted to compare teachers' answers with students' answers. This could be done either by adapting the MaTFAKi to ask students about what they think their mathematics teacher 'would do' in that situation or what they 'would like' their teachers to do. This would provide important information in terms of how students perceive their teachers' practices and to corroborate or confront teachers' actual practices.

Fourth, going back to my own experience in the first year of my PhD, I believe that now that more data is available, it is more likely that participatory research, focussing on the improvement of teachers' FA practices could be conducted. In this case, I would suggest starting with teachers' assessment methods due to its importance in the Brazilian context and the low performance of the items addressing this aspect in the MaTFAKi. Considering the importance of feedback to the implementation of FA, and teachers' divided performance that the results have shown, research focussing on teachers' views, the development of teachers' provision of feedback and encouraging feedback between students is of paramount importance.

Finally, considering that there was almost no variation in teachers' knowledge when comparing the results in terms of their level of education, years of experience and whether or not they have the teaching qualification in mathematics, further research is necessary to better understand what is shaping mathematics teachers' knowledge of FA.

11.5 Final remarks

In the work presented in this dissertation, I attempted to capture a picture of the FA knowledge of mathematics teachers in the FD of Brazil, by developing a structured instrument – the MaTFAKi.

There is more work needed to improve the MaTFAKi, but this is, to my knowledge, the first study focussing on mathematics teachers' knowledge of FA. The instrument here developed is the only one of its kind and provides a novel approach to research in this area.

Although the generalisation of results should be treated with caution, this study offered a valuable indication about the state of mathematics teachers' knowledge of FA in the Brazilian FD; information that was not available before.

As a result, this study adds to the mounting number of research studies on FA (ARG, 1999, 2002; Birenbaum, Kimron, & Shilton, 2011; Black et al., 2003; Black & Wiliam, 1998; Hodgen & Marshall, 2005; Laveault & Allal, 2016; O'Shea, 2015; Rakoczy et al., 2013; Sadler, 1989; Santos & Pinto, 2009; Wiliam, 1999). It extends such work by focusing on teacher knowledge, mathematics education and the use of an instrument for large scale data collection.

Again, I recognise that this was an ambitious study for the time frame and limited resources of a PhD study, but the results have shown that further discussion remains necessary, and this is one of the reasons why this study was valuable.

Looking back on the study, I am confident that my personal preferences did not diminish its value and that my familiarity with the context was actually beneficial to the results. In making these choices, I hoped to contribute to the current research debate on FA and provide effective and formative feedback to teachers. Giving this feedback to individual teachers from my own country regarding their current situation in terms of their knowledge of FA will in turn, I believe, better explain FA and help them effectively integrate it into their practice. Therefore, I would like to think that this study provides a starting point and a language for discussing reform, innovation, and change in FA and mathematics pedagogy in secondary schools not only in the FD but in Brazil as a whole.

REFERENCES

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New Jersey: Wiley-Interscience.

Albuquerque, L. C. (2012). *Avaliação da aprendizagem: concepções e práticas do professor de matemática dos anos finais do ensino fundamental* [Classroom-based assessment: conceptions and practices of upper-secondary mathematics teachers] (Unpublished master's thesis). Universidade de Brasília, Brasília, Brazil.

Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In *International conference on intelligent tutoring systems* (pp. 443–454).

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students.* Washington, DC.

Anderson, J. C., Gerbing, D. W., & Hunter, J. E. (1987). On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria. *Journal of Marketing Research*, *24*(4), 432–437.

Anthony, G., & Walshaw, M. (2002). Children's notions of commutativity: Do we need more than "turn-arounds"? *SET: Research Information for Teachers*(3), 8–11.

Arsham, H. (2014). *Questionnaire design and surveys sampling.* Retrieved 30^{th} September 2014, from `http://home.ubalt.edu/ntsbarsh/stat-data/Surveys.htm#rssm`

Askew, M. (2008). Mathematical discipline knowledge requirements for prospective primary teachers, and the structure and teaching approaches of programs designed to develop that knowledge. In P. Sullivan & T. Wood (Eds.), *The international handbook of mathematics teacher education* (Vol. 1, pp. 13–35). Rotterdam, The Netherlands: Sense Publishers.

Askew, S. (2000). *Feedback for learning*. Psychology Press.

Assessment Reform Group. (1999). *Assessment for learning: beyond the black box.* Cambridge, UK: University of Cambridge, School of Education.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles.* Cambridge, UK: University of Cambridge, Faculty of Education.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: what makes it special? *Journal of Teacher Education*, *59*(5), 389–407.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Barry, V. J. (2008). *Using descriptive feedback in a sixth grade mathematics classroom* (Unpublished master's thesis). University of Nebraska-Lincoln.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: issues and practice*, *17*(3), 37–45.

Bee, S. N. K., & Kaur, B. (2014). Using enhanced feedback to improve the learning of mathematics. *The mathematics Educator*, *15*(2), 101–119.

Bell, J. (2014). *Doing your research project: A guide for first-time researchers* (6th ed.). McGraw-Hill Education (UK).

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.

Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, *21*(1), 65–88.

BERA. (2011). *Ethical guidelines for educational research.* British Educational Research Association.

Bernardo, A. B. I. (1999). Overcoming obstacles to understanding and solving word problems in mathematics. *Educational Psychology*, *19*(2), 149–163.

Bicknell, B. (1999). The writing of explanations and justifications in mathematics: Differences and dilemmas. In *Making the difference: Proceedings of the* 22nd *annual conference of the Mathematics Education Research Group of Australasia* (pp. 75–83).

Birenbaum, M., Kimron, H., & Shilton, H. (2011). Nested contexts that shape assessment for learning: School-based professional learning community and classroom culture. *Studies in Educational Evaluation*, *37*(1), 35–48.

Bishop, A. J., Clements, K., Keitel, C., Kilpatrick, J., & Laborde, C. (2013). *International handbook of mathematics education*. Dordrecht: Springer Netherlands.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science*

Education, *21*(1), 49–97.

Black, P., & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: the science student's role in formative assessment. *School Science Review*, *83*(302), 43–49.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, *18*(4), 451–469.

Black, P., Harrison., C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Berkshire: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Black, P., & Wiliam, D. (2012). The reliability of assessments. In J. Gardner (Ed.), (pp. 243–263). London: Sage Publications Ltd.

Blaikie, N. (2000). *Designing social research: the logic of anticipation*. Cambridge: Polity Press.

Bleiler, S. K., Thompson, D. R., & Krajčevski, M. (2014). Providing written feedback on students' mathematical arguments: proof validations of prospective secondary mathematics teachers. *Journal of mathematics Teacher Education*, *17*(2), 105–127.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping-disaffection, polarisation and the construction of failure. *British Educational Research Journal*, *26*(5), 631–648.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, *22*(2), 151–167.

Boud, D., Cohen, R., & Sampson, J. (2001). *Peer learning in higher education: Learning from and with each other*. Routledge.

Bray, W. S. (2011). A collective case study of the influence of teachers' beliefs and knowledge on error-handling practices during class discussion of mathematics. *Journal for research in mathematics education*, *42*(1), 2–38.

Briggs, D. C., & Alonzo, A. C. (2012). The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science* (pp. 293–316). Sense Publishers.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, *11*(1), 33–63.

Bromme, R., & Tillema, H. (1995). Fusing experience and theory: The structure of professional knowledge. *Learning and instruction*, *5*(4), 261–67.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, *10*(2), 161–180.

Brookhart, S. M. (1998). Teaching about grading and communicating assessment results. In *The annual meeting of the national council on measurement in education.* San Diego, CA: ERIC.

Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, *18*(1), 5–13.

Brookhart, S. M. (2001). The standards and classroom assessment research. In *Proceedings of the Annual meeting of the american association of colleges for teacher education.* Dallas, TX: ERIC.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5–12.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Association for Supervision and Curriculum Development.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in mathematics*, *57*(2), 213–227.

Brown, G. T. L. (2003). *Teachers' conceptions of assessment* (Unpublished doctoral dissertation). University of Auckland.

Brown, G. T. L. (2008). Assessment literacy training and teachers' conceptions of assessment. In C. R. C. Rubie-Davies (Ed.), *Challenging thinking about teaching and learning* (pp. 285–302). Nova Science New York, NY.

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1995). *Cognitive psychology and instruction* (2nd ed.). Englewood Cliffs: Merrill.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1–14.

Cabrera, M. A. M., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and*

Assessment, *9*(1-2), 103–113.

Camargo, M. (2012). *Assessment practices in secondary-school mathematics teaching in Brazil* (Unpublished master's thesis). Faculty of Education, University of Cambridge, UK.

Camargo, M. (2015). Teachers' perceptions and use of assessment information: an exploratory study of mathematics teachers in Brazil. *International Journal for Research in mathematics Education*, *5*(1), 73–94.

Camargo, M., & Ruthven, K. (2014). Assessment practices in secondary-school mathematics teaching in Brazil. In S. Pope (Ed.), *Proceedings of the* 8^{th} *British congress of mathematics education 2014* (pp. 41–48).

Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 71–86). London: Sage Publications Ltd.

Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers. In *Annual meeting of the mid-western educational research association.* Columbus, OH.

Cardoso, E., & Souza, A. P. (2004). *The impact of cash transfers on child labor and school attendance in Brazil* (No. 04-W07). Nashville, TN: Department of Economics, Vanderbilt University.

Carr, D. (1992). Practical enquiry, values and the problem of educational theory. *Oxford Review of Education*, *18*(3), 241–251.

Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, *26*(4), 347–372.

Castelló, M., & Monereo, C. (2005). Students' note-taking as a knowledge-construction tool. *L1-Educational Studies in Language and Literature*, *5*(3), 265–285.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*(1), 143–159.

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*(3), 233–254.

Chan, P. E., Konrad, M., Gonzalez, V., Peters, M. T., & Ressa, V. A. (2014). The critical role of feedback in formative instructional practices. *Intervention in School and Clinic*, *50*(2), 96–104.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: constructs assessed and a meta-analysis of their criterion-related validities.

Personnel Psychology, *63*(1), 83–117.

Clariana, R. B. (1999). *Differential memory effects for immediate and delayed feedback: A delta rule explanation of feedback timing effects.* A Presentation at the Annual Convention of the Association for Educational Communications and Technology, Houston, OH: ERIC.

Clarke, D. J. (1992). Activating assessment alternatives in mathematics. *The Arithmetic Teacher*, *39*(6), 24–29.

Clarke, S., & Fisher, S. (2009). *Formative assessment in the secondary classroom*. Hodder.

Cobb, P., Wood, T., & Yackel, E. (1993). Discourse, mathematical thinking, and classroom practice. In E. Forman, N. Minick, & C. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 91–119). Oxford, UK: Oxford University Press.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational researcher*, *32*(6), 3–12.

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, *24*(1), 249–305.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London: Routledge.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 245–252).

Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: Motivational and performance consequences. *Journal of Educational psychology*, *76*(6), 1038–1050.

Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum Journal*, *16*(2), 137–151.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, *83*(1), 17–27.

Creswell, J. (2014). Survey designs. In *Educational research: planning, conducting and evaluating quantitative and qualitative research* (pp. 401–447). Harlow, Essex: Pearson.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review* *of Educational Research*, *58*(4), 438–481.

Dante, L. R. (2008). Livro didático de matemática: uso ou abuso? [mathematics textbooks: use or abuse?] *Em Aberto*, *16*(69), 83–97.

Darling-Hammond, L. (2015). *Powerful learning: What we know about teaching for understanding*. John Wiley & Sons.

Davis, G. E., & McGowen, M. A. (2007). Formative feedback and the mindful teaching of mathematics. *Australian Senior mathematics Journal*, *21*(1), 19–30.

Day, J. D., & Cordón, L. A. (1993). Static and dynamic measures of ability: An experimental comparison. *Journal of educational psychology*, *85*(1), 75–82.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438.

Dohrn, E., & Bryan, T. (1994). Attribution instruction. *Teaching Exceptional Children*, *26*(4), 61–63.

Donovan, D., & Loch, B. (2013). Closing the feedback loop: engaging students in large first-year mathematics test revision sessions using pen-enabled screens. *International Journal of Mathematical Education in Science and Technology*, *44*(1), 1–13.

Dudley, P., & Swaffield, S. (2008). Understanding and using assessment data. In S. Swaffield (Ed.), *Unlocking assessment* (pp. 105–120). Taylor & Francis.

Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, *41*(10), 1040–1048.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Earl, L. M., & Katz, M. S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship & Youth.

Earl, L. M., & Katz, M. S. (2008). Getting the core of the learning: using assessment for self-monitoring and self-regulation. In S. Swaffield (Ed.), *Unlocking assessment: understanding for reflection and application* (pp. 90–104). Oxon: Routledge.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, *77*(2), 162–173.

Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In P. C. Claire Wyatt-Smith

Val Klenowski (Ed.), *Designing assessment for quality learning* (pp. 91–107).
English, L. D., & Kirshner, D. (2015). *Handbook of international research in mathematics education*. New York: Routledge.
Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2010). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*(2), 187–201.
Fagerland, M. W., & Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary clinical trials*, *30*(5), 490–496.
Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of research in education*, *20*(1), 3–56.
Fernandes, R. (2007). *Índice de desenvolvimento da educação básica (IDEB). Texto para discussão n. 26* [Index of basic education development]. Retrieved 16^{th} March 2016, from `http://www.publicacoes.inep.gov.br/portal/download/503`
Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.
Fink, A. (2008). *How to conduct surveys: a step-by-step guide* (3rd ed.). Los Angeles: Sage Publications, Inc.
Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational researcher*, *18*(9), 27–32.
Frey, N., & Fisher, D. (2011). *The formative assessment action plan: Practical steps to more successful teaching and learning*. ASCD.
Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: an introduction* (2nd ed.). London: SAGE Publications Ltd.
Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, *49*(9), 1181–1210.
Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching a meta-analysis. *Review of educational research*, *82*(3), 300–329.
Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13.

Garcia, T. (1995). The role of motivational strategies in self-regulated learning. *New directions for teaching and learning*, *1995*(63), 29–42.

Gardner, J. (2012). Quality assessment practice. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 103–121). London: Sage Publications Ltd.

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at horn's parallel analysis with ordinal variables. *Psychological methods*, *18*(4), 454–474.

Gatti, B. A. (2010). Formação de professores no Brasil: características e problemas [Teachers' education in Brazil: characteristics and problems]. *Educ. Soc.*, *31*(113), 1355–1379.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of educational psychology*, *76*(4), 569–582.

Good, R. (2011). Formative use of assessment information: It's a process, so let's say what we mean. *Practical Assessment, Research & Evaluation*, *16*(3), 2–6.

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, *23*(5), 531–557.

Governo do Distrito Federal, Secretaria de Estado de Educação do Distrito Federal, & Subsecretaria de Educação Básica. (2014). *Diretrizes de avaliação educacional* [Guidelines of educational assessment]. Brasília: Governo do Distrito Federal.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through pre-service teacher education. *Teaching and Teacher Education*, *21*(6), 607–621.

Grossman, P. L. (1995). Teachers' knowledge. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (Vol. 2, pp. 20–24).

Guo, S. W., & Thompson, E. A. (1989). *Analysis of sparse contingency tables: Monte Carlo estimation of exact p-values* (Technical Report No. 187). Seattle, Washington: University of Washington.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Oxon, UK: Routledge. Kindle Edition.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, *35*(2), 213–224.

Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: reconceptualising assessment for learning. *Oxford Review of Education*, *39*(2), 229–246.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, *34*(3), 269–290.

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 87–102). London: Sage Publications Ltd.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J., & Jaeger, R. (1998). Assessment and classroom learning: A deductive approach. *Assessment in Education*, *5*(1), 111–122.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81–112.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of educational psychology*, *87*(1), 18–32.

Heinze, A. (2005). Mistake-handling activities in the mathematics classroom. In *Proceedings of the* 29^{th} *conference of the international group for the psychology of mathematics education* (Vol. 3, pp. 105–112).

Heitink, M., der Kleij, F. V., Veldkamp, B., Schildkamp, K., & Kippers, W. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50–62.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177–189.

Heritage, M. (2007). Formative assessment: what teachers need to know and do? *The Phi Delta Kappan*, *89*(2), 140–145.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, *28*(3), 24–31.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing & health*, *31*(2), 180–191.

Hettmansperger, T. P., & McKean, J. W. (2010). *Robust nonparametric statistical methods*. CRC Press.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies* *in Higher Education*, *27*(1), 53–64.

Hodgen, J., & Marshall, B. (2005). Assessment for learning in english and mathematics: A comparison. *Curriculum journal*, *16*(2), 153–176.

Hodgen, J., & Wiliam, D. (2006). *mathematics inside the black box: assessment for learning in the mathematics classroom*. London: GL Assessment.

Hoffmann, J. (2005). *O jogo do contrário em avaliação* [Playing the opposite game in assessment]. Mediação.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153–166.

Hollander, J. A. (2002). Learning to discuss: Strategies for improving the quality of class discussion. *Teaching Sociology*, *30*(3), 317–327.

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods*. John Wiley & Sons.

Hutchenson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage.

Jorro, A. (2000). *L'enseignant et l'évaluation: des gestes évaluatifs en question*. Bruxelles: De Boeck Supérieur.

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, *23*, 33–42.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger Publishers.

Kay, R., & Knaack, L. (2009). Exploring the use of audience response systems in secondary school science classrooms. *Journal of Science Education and Technology*, *18*(5), 382–392.

Kearney, W. S., Webb, M., Goldhorn, J., & Peters, M. L. (2013). Examining the impact of critical feedback on learner engagement in secondary mathematics classrooms: A multi-level analysis. *AASA Journal of Scholarship and Practice*, *10*(1), 23–38.

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, *15*(3), 261–266.

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*(2), 235–245.

Kershaw, I. (1993). *Ohio vocational education teachers' perceived use of student assessment information in educational decision making* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Key, J. P. (1997). *Chi square.* Retrieved from `http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm`

Khisty, L. L., & Chval, K. B. (2002). Pedagogic discourse and equity in mathematics: When teachers' talk matters. *mathematics Education Research Journal*, *14*(3), 154–168.

Kiemer, K., Gröschner, A., Pehmer, A. K., & Seidel, T. (2015). Effects of a classroom discourse intervention on teachers' practice and students' motivation to learn mathematics and science. *Learning and Instruction*, *35*, 94–103.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37.

Klassen, R., Durksen, T., Rowett, E., & Patterson, F. (2014). Applicant reactions to a situational judgment test used for selection into initial teacher training. *International Journal of Educational Psychology*, *3*(2), 104–124.

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, *2*(2), 145–163.

Kline, P. (1986). *A handbook of test construction: introduction to psychometric design*. London: Methuen & Co. Ltd.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, *7*(3), 67–72.

Koh, K., Hadden, J., Parks, C., Sanden, L., Monaghan, M., Gallant, A., & LaFrance, M. (2015). Building teachers' capacity in authentic assessment and assessment for learning. In P. Preciado Babb, M. Takeuchi, & J. Lock (Eds.), *Proceedings of the ideas: Designing responsive pedagogy conference* (pp. 43–52). Calgary, Canada: Werklund School of Education, University of Calgary.

Kovalainen, M., & Kumpulainen, K. (2005). The discursive practice of participation in an elementary classroom community. *Instructional Science*, *33*(3), 213–250.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173–194.

Latham, G. P., & Locke, E. A. (1979). Goal setting – a motivational technique that works. *Organizational Dynamics*, *8*(2), 68–80.

Laveault, D. (2013). Summative assessment for learning: How it may impact task design. In *15th conference of the european association for research on learning and instruction.* Munich, Germany.

Laveault, D. (2016). Building capacity: professional development and collaborative learning about assessment. In D. Levault & L. Allal (Eds.), *Assessment for learning: meeting the challenge of implementation* (pp. 131–143).

Laveault, D., & Allal, L. (2016). Implementing assessment for learning: Theoretical and practical issues. In *Assessment for learning: Meeting the challenge of implementation* (pp. 1–18).

Leinhardt, G., Young, K. M., & Merriman, J. (1995). Integrating professional knowledge: The theory of practice and the practice of theory. *Learning and instruction*, *5*(4), 401–408.

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology*, *15*(4), 319–333.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, *33*(3), 263–275.

London, M., & Sessa, V. I. (2006). Group feedback for continuous learning. *Human Resource Development Review*, *5*(3), 303–329.

Looney, J. W. (2011). *Integrating formative and summative assessment: progress toward a seamless system?* OECD Education Working Paper No. 58. Retrieved January 3rd, from `http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=edu/wkp(2011)4&doclanguage=en`

Loughran, J. (2012). *What expert teachers do: Enhancing professional knowledge for classroom practice*. Routledge.

Lyon, C., Miller, S., & Wylie, E. C. (2011). A validity argument for the assessment practices survey: A measure of teachers' enactment of formative assessment. In *Annual meeting of the american educational research association.* New Orleans, LA.

Lyons, P., & Doueck, H. J. (2010). *The dissertation: from beginning to end*. Oxford:

Oxford University Press.

Lysaght, Z., & O'Leary, M. (2013). An instrument to audit teachers' use of assessment for learning. *Irish Educational Studies*, *32*(2), 217–232.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in china and the united states*. Lawrence Erlbaum Associates Mahwah, NJ.

MacGregor, M. (2002). Using words to explain mathematical ideas. *The Australian Journal of Language and Literacy*, *25*(1), 78–88.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, *20*(5), 523–535.

Mansell, W., James, M., & the Assessment Reform Group. (2009). *Assessment in Schools. Fit for Purpose? A Commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council, Teaching and Learning Research Programme.

Marsh, H. W. (1990). Influences of internal and external frames of reference on the formation of Math and English self-concepts. *Journal of Educational Psychology*, *82*(1), 107–116.

Martiniello, M. (2008). Language and the performance of english-language learners in math word problems. *Harvard Educational Review*, *78*(2), 333–368.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us.* Retrieved 25th October 2014, from `http://dwb.unl.edu/Edit/MB/MasonBruning.html`

Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower-and higher-achieving urban schools. *The Elementary School Journal*, *103*(1), 3–25.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: a clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730–740.

McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, *33*(5), 515–525.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation online*, *7*(8).

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. 3–16). London: Sage Publications Ltd.

Melo, R. d. M. S., & Duarte, G. B. (2010). Impacto do programa Bolsa Família sobre a frequencia escolar: o caso da agricultura familiar no Nordeste do Brasil [The impact of the bolsa familia programme on school attendance: the case of family agriculture in the northeast of Brazil]. *Revista de Economia e Sociologia Rural*, *48*(3), 635–657.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon, UK: Multilingual Matters.

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British journal of educational psychology*, *80*(1), 1–14.

Mertler, C. A. (2003). Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference? In *Annual meeting of the mid-western educational research association.* Columbus, OH.

Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge and applications of classroom assessment concepts: development of the assessment literacy inventory. In *Annual meeting of the american educational research associations.* Montréal, Quebec, Canada.

Michalewicz, Z., & Fogel, D. B. (2013). *How to solve it: modern heuristics*. Springer Science & Business Media.

Millett, A., Askew, M., & Brown, M. (2004). The impact of the national numeracy strategy in year 4 (ii): Teaching. *Research in mathematics Education*, *6*(1), 191–205.

Ministério da Educação, Secretaria de Educação Básica, & Diretoria de Currículos e Educação Integral. (2013). *Diretrizes curriculares nacionais da educação básica* [National curriculum guidelines for basic education]. Brasília: Ministério da Educação.

Moore, W., & Esselman, M. (1992). Teacher efficacy, power, school climate and achievement: A desegregating district's experience. In *annual meeting of the American Educational Research Association, San Francisco.*

Moreira, C. (2013). *Distorção idade-série na educação básica* [Age-grade distortion in Basic Education]. Retrieved 15th December 2014, from http://cmoreira2.jusbrasil.com.br/artigos/111821615/

distorcao-idade-serie-na-educacao-basica

Morris, M. (1998). Ethical challenges. *The American Journal of Evaluation*, *19*(3), 381–382.

Morris, M. (2000). Ethical challenges. *The American Journal of Evaluation*, *21*(1), 119–121.

Morrison, K. (1993). *Planning and accomplishing school-centred evaluation*. Peter Francis.

Moschkovich, J. (1999). Supporting the participation of english language learners in mathematical discussions. *For the learning of mathematics*, *19*(1), 11–19.

Moschkovich, J. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical thinking and learning*, *4*(2-3), 189–212.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*(6), 640–647.

Muijs, D., & Reynolds, D. (2010). *Effective teaching: Evidence and practice*. Sage.

Multon, K. D. (2010). Test–retest reliability. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1496–1499). SAGE Publications, Inc.

Muniz, C. A. (2009). A produção de notações matemáticas e seu significado. [The production and meanings of mathematical notations.] In M. H. Fávero & C. da Cunha. (Eds.), *Psicologia do conhecimento: o diálogo entre as ciências e a cidadania.* (pp. 115–143). Brasília: Unesco and UnB.

Nathan, M. J., & Knuth, E. J. (2003). A study of whole classroom mathematical discourse and teacher change. *Cognition and instruction*, *21*(2), 175–207.

National Council of Teachers of mathematics. (2000). *Standards for school mathematics.* Reston, VA.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155–175.

Newby, L., & Winterbottom, M. (2011). Can research homework provide a vehicle for assessment for learning in science lessons? *Educational Review*, *63*(3), 275–290.

Newcombe, R. G. (1998a). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in medicine*, *17*(8), 873–890.

Newcombe, R. G. (1998b). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in medicine*, *17*(8), 857–872.

Newman, J. (2016). *Differentiated feedback: Responding to students' individual needs.* Retrieved 30th August 2016, from `http://hdl.handle.net/1807/72255`

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199–218.

Norris, S. P. (2000). The pale of consideration when seeking sources of teaching expertise. *American Journal of Education*, *108*(3), 167–195.

Núñez, J. C., Suárez, N., Rosário, P., Vallejo, G., Cerezo, R., & Valle, A. (2015). Teachers' feedback on homework, homework-related behaviors, and academic achievement. *The Journal of Educational Research*, *108*(3), 204–216.

Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers Pédagogiques*, *280*, 47–64.

Ní Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, *18*(2), 219–233.

OECD. (2014). *Pisa 2012 results: What students know and can do – student performance in mathematics, reading and science* (Vol. 1).

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*, *47*(4), 422–453.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: St. Martin Press.

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, *27*(4), 309–323.

O'Shea, A. (2015). *Exploring the black box: a multi-case study of assessment for learning in mathematics and the development of autonomy with 9-10 year old children.* (Unpublished doctoral dissertation). University of Cambridge.

Owen, L. (2016). The impact of feedback as formative assessment on student performance. *International Journal of Teaching and Learning in Higher Education*, *28*(2), 168–175.

Pacheco, M. M. D. R. (2007). *Concepções e práticas avaliativas nos cursos de licenciatura* [Concepts and assessment practices in undergraduate courses] (Unpublished doctoral dissertation). Pontifícia Universidade Católica de São Paulo, São Paulo.

Pellegrino, J. W. (2006). *Rethinking and redesigning curriculum, instruction and assessment: What contemporary research and theory suggests* (Tech. Rep.). Washington, DC: National Center on Education and the Economy for the *New* Commission on the Skills of the American Workforce.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, *6*(1), 21–27.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: what do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: learning, adjustment, and achievement* (pp. 53–68). London: Academic Press Limited.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment*, *11*(1), 1–16.

Polya, G. (2014). *How to solve it: A new aspect of mathematical method*. Princeton university press.

Popham, W. J. (2008). *Transformative assessment.* Alexandria: Association for Supervision & Curriculum Development.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4–11.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: all that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, *35*(3), 277–289.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, *9*(3), 241–258.

Radhakrishna, R. B. (2007). Tips for developing and testing questionnaires/instruments. *Journal of Extension*, *45*(1), 1–4.

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, *27*, 63–73.

Randel, B., Beesley, A. D., Apthorp, H., Clark, T. F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). *Classroom assessment for student learning: The* *impact on elementary school mathematics in the central region* (Tech. Rep. No. NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance: Institute of Education Sciences, U.S. Department of Education.

Randel, B., & Clark, T. (2013). Measuring classroom assessment practices. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 145–163). London: Sage Publications Ltd.

Reeve, B. B., & Mâsse, L. C. (2004). Item response theory modeling for questionnaire evaluation. In S. Presser et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 247–273). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, *41*(2), 301–315.

Robson, C. (2011). *Real world research* (3rd ed.). Chichester: John Wiley & Sons.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, *17*(1), 1–24.

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education/Revue canadienne de l'education*, *17*(1), 51–65.

Rowe, K. (2004). The importance of teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision–implications of findings from the international and Australian evidence-based research. Melbourne, Australia.

Ruiz-Primo, M. A., Furtak, E. M., Ayala, C., Yin, Y., & Shavelson, R. J. (2010). Formative assessment, motivation, and science learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 139–158). New York: Routledge.

Rural Alaska Principal Preparation and Support Project. (n.d.). *Actionable feedback strategies for the classroom.* Retrieved 22nd October 2014, from http://rapps.pbworks.com/f/MP+Handouts+Formative+Feedback+6-3-10.pdf

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, *28*(2), 147–164.

Rust, J., & Golombok, S. (2009). *Modern psychometrics: the science of psychological assessment* (3rd ed.). Hove, East Sussex: Routledge.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, *17*(4), 688–690.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119–144.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*(5), 535–550.

Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods, and techniques. In N. R. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work, & organizational psychology* (Vol. 1). London: Sage.

Salili, F., & Hau, K. T. (1994). The effect of teachers' evaluative feedback on chinese students' perception of ability: A cultural and situational analysis. *Educational Studies*, *20*(2), 223–236.

Santagata, R. (2005). Practices and beliefs in mistake-handling activities: A video study of Italian and US mathematics lessons. *Teaching and Teacher Education*, *21*(5), 491–508.

Santos, L., & Pinto, J. (2009). Lights and shadows of feedback in mathematics learning. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Eds.), *Proceedings of the* 33*rd* *conference of the international group for the psychology of mathematics education* (Vol. 1, pp. 49–56). Thessaloniki, Greece: PME.

Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school english language arts and mathematics. *Journal of Multidisciplinary Evaluation*, *8*(17), 1–24.

Schön, D. A. (1992). The theory of inquiry: Dewey's legacy to education. *Curriculum inquiry*, *22*(2), 119–139.

Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). *Conducting research surveys via e-mail and the web*. Rand Corporation.

Scott, S., Webber, C. F., Aitken, N., & Lupart, J. (2011). Developing teachers' knowledge, beliefs, and expertise: Findings from the Alberta Student Assessment Study. *The Educational Forum*, *75*(2), 96–113.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, UK: Cambridge University Press.

Scriven, M. (1967). The methodology of evaluation. In M. S. Ralph W. Tyler Robert M. Gagné (Ed.), *Perspectives of curriculum evaluation* (pp. 39–93). Chicago, IL: Rand McNally.

Sebba, J., Crick, R. D., Yu, G., Lawson, H., Harlen, W., & Durant, K. (2008). *Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment* (Technical Report No. 1614). London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Secretaria de Educação Fundamental. (1998). *Parâmetros curriculares nacionais: matemática* [National curriculum parameters]. Brasília, DF: Ministério da Educação.

Segers, M., Dochy, F., & Gijbels, D. (2010). Impact of assessment on students' learning strategies and implications for judging assessment quality. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., Vol. 3, pp. 196–201). Elsevier Ltd.

Shepard, L. A. (2002). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 1066–1101). Washington: American Educational Research Association.

Shepard, L. A. (2013). Foreword. In J. H. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. xix–xxii). London: Sage Publications Ltd.

Shulman, L. S. (1999). Knowledge and teaching: Foundations of the new reform. In J. Leach & B. Moon (Eds.), *Learners and pedagogy* (pp. 61–77). Paul Chapman London in association with The Open University.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371–391.

Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationship of perceived autonomy support and structure in the prediction of self-regulated learning. *British Journal of Educational Psychology*, *79*(1), 57–68.

Simmons, M., & Cope, P. (1993). Angle and rotation: Effects of different types of feedback on the quality of response. *Educational Studies in mathematics*, *24*(2), 163–176.

Simões, A. A. (2012). *The contribution of Bolsa Família to the educational achievement of economically disadvantaged children in Brazil.* (Unpublished doctoral dissertation). University of Sussex, Sussex, UK.

Sirkin, R. M. (2006). *Statistics for the social sciences*. London: SAGE.

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading.

Educational Measurement: Issues and Practice, *22*(4), 26–33.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2005). *Comparing check-all and forced-choice question formats in web surveys: the role of satisficing, depth of processing, and acquiescence in explaining differences* (Tech. Rep. No. 05-029). Pullman, Washington: Social & Economic Sciences Research Center.

Soares, J. F., & Xavier, F. P. (2013). Pressupostos educacionais e estatísticos do IDEB [Educational and statistical assumptions of IDEB]. *Educação & Sociedade*, *34*(124), 903–923.

Society, B. P. (2013). *Ethics guidelines for internet-mediated research. INF206/1.2013.* Leicester: Author. Retrieved 6^{th} February 2016, from `www.bps.org.uk/publications/policy-andguidelines/research-guidelines-policydocuments/research-guidelines-poli`

Son, J. W. (2013). How preservice teachers interpret and respond to student errors: ratio and proportion in similar rectangles. *Educational studies in mathematics*, *84*(1), 49–70.

Steinberg, R. M., Empson, S. B., & Carpenter, T. P. (2004). Inquiry into children's mathematical thinking as a means to teacher change. *Journal of mathematics Teacher Education*, *7*(3), 237–267.

Stiggins, R. (1991). Assessment literacy. *The Phi Delta Kappan*, *72*(7), 534–539.

Stiggins, R. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, *18*(1), 23–27.

Stiggins, R. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, *20*(3), 5–15.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *The Phi Delta Kappan*, *83*(10), 758–765.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324–328.

Stiggins, R. (2007). Conquering the formative assessment frontier. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 8–28). New York: Teachers College Press.

Stiggins, R. (2009). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment: learning, adjustment, and achievement* (pp. 233–250). New York: Routledge.

Stiggins, R. (2011). *An introduction to student-involved assessment for learning* (6th ed.). Boston, MA: Pearson Education, Inc.

Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 233–242).

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: a practical guide to their development and use* (2nd ed.). Oxford University Press, USA.

Suskie, L. (2002). Fair assessment practices: giving students equitable opportunities to demonstrate learning. *Adventures in Assessment*, *14*, 5–10.

Suskie, L. (2009). *Assessing student learning: a common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 433–449.

Sztajn, P. (2000). Sem óculos ou mau humor: somos professores de matemática [No glasses or bad mood: we are mathematics teachers]. In V. M. Candau (Ed.), *Reinventar a escola* (pp. 221–237). Petrópolis, RJ: Vozes.

Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, *28*(5), 549–565.

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478.

Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, *40*(3), 435–456.

Veldhuis, M., & van den Heuvel-Panhuizen, M. (2014). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE*, *9*(1), e86817.

Villas Boas, B. M. d. F. (2008). Auto-avaliação, avaliação e portfólio [Self-assessment, assessment and portfolios]. In *Virando a escola do avesso por meio da avaliação* (pp. 81–90). Campinas: Papirus.

Villas Boas, B. M. d. F. (2011). Compreendendo a avaliação formativa [Understanding formative assessment]. In *Avaliação formativa: práticas inovadoras* (pp. 30–42). Campinas: Papirus.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, *30*(3), 749–770.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516–551.

Watkins, C. (2000). Feedback between teachers. In S. Askew (Ed.), *Feedback for learning* (pp. 65–80). Psychology Press.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, *79*(1), 1–28.

Weeden, P., & Winter, J. (1999). *Learners' expectations of assessment for learning nationally.* University of Bristol, Graduate School of Education, CLIO Centre for Assessment Studies. Report for the Qualifications and Curriculum Authority (QCA).

Weekley, J. A., & Ployhart, R. E. (2013). *Situational judgment tests: Theory, measurement, and application*. Psychology Press.

Welch, B. L. (1947). The generalization of student's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

Wells, G., & Arauz, R. M. (2006). Dialogue in the classroom. *The journal of the learning sciences*, *15*(3), 379–428.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*(3), 188–202.

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, *21*(3), 291–309.

White, D. Y. (2003). Promoting productive mathematical classroom discourse with diverse students. *The Journal of Mathematical Behavior*, *22*(1), 37–53.

Whitington, V., Glover, A., & Harley, F. (2004). Preservice early childhood students' perceptions of written feedback on their essays. *Early Child Development and Care*, *174*(4), 321–337.

Wiliam, D. (1999). Formative assessment in mathematics part 2: feedback. *Equals: mathematics and Special Educational Needs*, *5*(3), 8–11.

Wiliam, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 183–204).

Wiliam, D. (2011). *Embedded formative assessment.* Bloomington, IN: Solution Tree Press [Kindle version].

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices*

From the Middle, *21*(2), 15–20.

Wiliam, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*(5), 537–548.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Hawker Brownlow Education.

Woodward, J., & Irwin, K. (2005). Language appropriate for the New Zealand numeracy project. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, & R. Pierce (Eds.), *Building connection: Theory, research and practice (proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 799–806). Sydney: MERGA.

Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, *10*(3).

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for research in mathematics education*, *27*(4), 458–477.

Yin, Y. (2005). *The influence of formative assessments on student motivation, achievement, and conceptual change* (Unpublished doctoral dissertation). Stanford University.

Yin, Y., Tomita, M. K., & Shavelson, R. J. (2014). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and achievement in science. *International Journal of Science Education*, *36*(4), 531–552.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education*, *45*(4), 477–501.

Zhang, Z., & Burry-Stock, J. A. (1994). *Assessment practices inventory*. Tuscaloosa, AL: The University of Alabama.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29.

APPENDICES

A Documents

A.1 Validation of the DoK

Dear Researcher,

As previously agreed, I set out below six Domains of Knowledge (and a brief explanation) that mathematics teachers in the FD should have in relation to formative assessment.

I. Knowledge of articulating clear learning intentions and success criteria (KLI) One of the first principles of formative assessment is clarifying and sharing learning intentions and success criteria (Black & Wiliam, 1998; Wiliam, 2011). Therefore, teachers should be able to articulate what is expected for students to learn in clear, attainable and assessable ways (Brookhart, 2011). Clear, so students can understand what is expected from them. Attainable, so students are able to achieve them. Assessable, so that both teachers and students will know whether and to what extent they have been achieved. Therefore, they are crucial to answer the question *where am I going?* in the feedback cycle.
II. Knowledge of designing or choosing assessment methods to collect evidence of students' learning (KAM) Teachers should understand the purposes and uses of the range of available assessment options, being aware that different methods can be incompatible with certain instructional goals and may impact quite differently on their teaching (AFT, NCME, & NEA, 1990).

For formative assessment, teachers should design or choose methods from which both teachers and students can use results (Black & Wiliam, 1998; McMillan, 2000; Stiggins, 2009). In essence, a method constructed to be used formatively will have the main purpose of giving feedback to students and will be based on a narrow range of learning intentions (Brookhart, 2011). It is through the methods chosen or designed that teachers will be able to understand where students are in their learning, and be able to give them feedback.

III. Knowledge of interpreting evidence of students' learning (KIE)

Knowing how to draw inferences from students' responses (be they oral or written) is crucial to the effectiveness of formative assessment and consequently to giving feedback to students (Black & Wiliam, 1998; Hattie & Timperley, 2007).

Essentially, teachers should be able to identify students' current mathematical understanding so that they can modify instruction to facilitate improvement. The analysis of students' responses can be either on formal (usually written) assessment, which usually will take place after the lesson and therefore the teacher will have more time to think and closely examine it; or informal (usually through observations or classroom discussions), in which the teacher will have to make inferences (and interventions) on a moment-by-moment basis.

In both cases, teachers should know how to translate their analyses into feedback to students – linked to the learning intentions and the success criteria (KLI above) – that can be used by students to further their learning. That is, to help them to answer the question *how am I going?* in the feedback cycle.

IV. Knowledge of providing effective feedback (KEF)

Effective teacher feedback on student work should be descriptive and should comment on the work itself and the process used to do the work (Hattie & Timperley, 2007). It should give students information about their work set against the criteria for good work that were articulated as part of their learning intentions and shared with students (KLI above). Effective feedback is elaborated but not too complex, is specific to the work, avoids general praise, and is different for different learners (Shute, 2008; Stiggins, 2009). Effective feedback is at an intermediate level of generality (Brookhart, 2008; Kluger & DeNisi, 1996) so that students can identify specific improvements that are needed but not so specific that the work is already done for them.

V. Knowledge of closing the feedback loop (KCL)

If the main goal of giving feedback to students is to help them in the improvement of their learning, students should be given the opportunity to act upon the feedback. That is, the only way to know whether the improvement of learning is a result from feedback is by giving opportunities to students to make some kind of response to complete the feedback loop (Sadler, 2010). These opportunities should be given while the assignment or learning task is still relevant to the student (Brookhart, 2008; Hattie & Timperley, 2007).

VI. Knowledge of helping students use assessment information (KHS)

A major contribution of the formative assessment literature is that the effects of assessment on students are powerful (Black & Wiliam, 1998). Therefore, it is one of the teacher's roles to help students to use assessment *for* learning. Giving effective and useful feedback, as well as encouraging self- and peer-assessment, play an essential role in this matter.

Please rate the importance of each statement.

	Not important at all	Somewhat important	Important	Very important
KLI	○	○	○	○
KAM	○	○	○	○
KIE	○	○	○	○
KEF	○	○	○	○
KCL	○	○	○	○
KHS	○	○	○	○

Please indicate whether you felt, based on your professional expertise, that any aspects of formative assessment were missing, or whether you would like to rephrase or exclude any of the statements.

A.2 Alignment between questions and DoK

Dear Researcher,

As agreed, please use the list of Domains of Knowledge sent to you previously and the questionnaire attached to this email, to rate how highly you feel the alignment is between the question and the domain which that specific question is supposed to measure.

Question	Domain	Very low	Low	Moderate	High	Very high
1	KIE	○	○	○	○	○
2	KEF	○	○	○	○	○
3	KLI	○	○	○	○	○
4	KCL	○	○	○	○	○
5	KAM	○	○	○	○	○
6	KIE	○	○	○	○	○
7	KLI	○	○	○	○	○
8	KHS	○	○	○	○	○
9	KEF	○	○	○	○	○
10	KCL	○	○	○	○	○
11	KAM	○	○	○	○	○
12	KHS	○	○	○	○	○
13	KIE	○	○	○	○	○
14	KHS	○	○	○	○	○
15	KAM	○	○	○	○	○
16	KLI	○	○	○	○	○
17	KAM	○	○	○	○	○
18	KHS	○	○	○	○	○
19	KEF	○	○	○	○	○
20	KCL	○	○	○	○	○
21	KLI	○	○	○	○	○
22	KIE	○	○	○	○	○
23	KEF	○	○	○	○	○
24	KCL	○	○	○	○	○

Thank you for your participation.

B Questionnaire versions - English

In this appendix, I include all English versions of the questionnaire. Table 11.2 presents the question equivalence among the versions. In the "main survey", most of the order of the response options is also different from the other versions.

Table 11.2: Correspondence among the different versions.

Main survey	Field test	Second validation	First validation
1	1	1	1
2	2	2	2
	3	3	3
3	4	4	4
4	5	17	11
5	6	18	*
6	7	19	
7	8	20	
8	9	9	6
9	10	10	7
	11	11	8
	12	12	
10	13	5	
	14	6	
11	15	7	
12	16	8	
13	17	21	12
14	18	22	13
15	19	23	
16	20	24	
17	21	13	

Continued on next page

Table 11.2 – continued from previous page.

Main survey	Field test	Second validation	First validation
18	22	14	9
19	23	15	10
20	24	16	
*Question 5 was deleted after the 'First validation' (See page 119).			

B.1 First version with 13 questions

Instructions

1. All answers are anonymous. No type of identification is associated with the submitted responses.

2. This is a special version for validation. Therefore, although the questions are asking you to choose *the best* option (because this is how the teachers will be asked), in this version I ask you to drag the options (A, B, C and D) and *put them in order*: from the most adequate to the least adequate, in relation to the situation presented.

3. In some cases, the difference between options is very subtle. However, this was done on purpose.

4. It is not possible to return to a question once you have moved to the next page.

5. This version is comprised of 13 questions embedded in 6 scenarios. It takes approximately 30 minutes to answer them all.

6. The final version will be comprised of multiple choice questions in which the teachers will be required to choose ***only*** one option - the one they judge to be *the best*.

Thank you very much for your participation.
Looking forward to seeing you on Tuesday, the 16^{th}.
Regards, Mel.

Scenario 1

In the last lesson, Mr. Smith taught his students how to solve inequalities. However, in the assigned homework, he observed that the majority of the class had trouble solving them. See some examples of how they were solving them:

1. $17 - 2x > 25$

$$2x - 17 > -25$$
$$2x > -25 + 17$$
$$2x > -8$$
$$x > -4$$

2. $23 + x > 3x - 9$

$$x - 3x > -9 - 23$$
$$-2x > -32 \; (-1)$$
$$2x > 32$$
$$x > \frac{32}{2}$$
$$x > 16$$

Q1. What is the best interpretation that can be made in relation to these patterns and the students' understanding of the subject?

(a) These students understand all the steps to solve inequalities, but they had a very simple mistake in the end.

(b) These students are solving as equations, because they did all the steps correctly, but they did not take the $>$ sign into account.

(c) These students do not know how to solve inequalities, because they did not find the correct results in any of the exercises.

(d) These students did not understand anything that the teacher taught about inequalities because they should have kept the x in the right hand side of the $>$ instead of bringing it to the left side.

Q2. Since the majority of students were having problems, what would be the best strategy to remedy these?

(a) Give written feedback to every student who had not been successful with the homework, and give them another opportunity to try to solve it.

(b) Re-teach that specific content to the whole group in the next lesson, and give the students another opportunity to practice them.
(c) Re-teach the content to the whole group in the next lesson.
(d) Return the homework assignment to students and give them the opportunity to re-do the exercise in class, asking for the teacher's help when they deem necessary.

Mr Smith was not very sure about what to do, but he decided to re-teach the content to the whole group.

Q3. What would be the best way to make the aim clear to students when he starts this conversation in the next lesson?

(a) "We will need to go over the homework assignment because many of you had trouble doing it and I want to give everyone the opportunity to learn what was taught in the last lesson."
(b) "Today, I will show you all again how to solve inequalities because many of you did it wrongly in your homework and I want you all doing it correctly."
(c) "In our lesson today, I will correct the homework assignment in the board because the majority of you got it wrong, which shows me that I need to re-teach it."
(d) "We are going to start our lesson by solving some inequalities because I noticed that some students had had trouble with the homework and I want everyone to learn how to solve inequalities."

Q4. Once the aim is clear, what would be the best thing for Mr Smith to do next?

(a) Repeat previous explanation and pass back the homework papers so students can try it again.
(b) Explain to students what kind of mistake they were making and pass back the homework papers so they can try it again.
(c) Pass back the homework to students and present the correct solution to the inequalities on the board, explaining each type of mistake displayed by students.
(d) Call one student who he knows had done it successfully to come to the board to demonstrate while the others solve it in their notebook.

Scenario 2

Mrs. Andrews is used to giving written feedback to each student on almost all the test questions that she assigns. In the last test she gave, she decided to change strategies, because she wanted to give her students the opportunity to discuss their ideas in pairs and try to solve the questions together.
When she returned the test to her students, she said:

> "This is the test I gave you in the last lesson. I will return your script and you will see that when you got a question right, I indicated it. When you didn't get it right, I did not write anything.
>
> What I want you to do is choose a friend, sit together and solve each question that you didn't get right. At the end of the lesson, I will collect your work and take it home to have a look again"

Working together like this was new to the students, and Mrs. Andrews observed that, instead of discussing the questions, the students were copying the correct answers from their colleagues.

Later, she told the other teachers what had happened so they could give her some advice. Although she had managed to work around the unexpected situation, she still wanted to try this strategy again, at another opportunity.

Q5. What would be the best advice to give to Mrs. Andrews?

(a) "The best way to do this is by putting together those students that made the same mistake on a particular question. By doing this, they will not copy each others' answers, they will have to discuss and solve the questions together."

(b) "I believe that you need to remind them about this experience and explain that they did not do what you asked them to do. Make clear to them that you will try again, but that will be their last chance, and the next time they will not have another opportunity to revise their test."

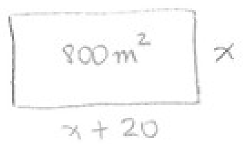
(c) "The most important part is to make clear to students what your intention with the activity is. That you are giving them a second opportunity of learning, and that you are putting them in pairs so they can discuss and help each other to reflect about their mistakes."

(d) “Working in pairs or groups is not easy. You need to give some time for your students to learn and get used to it. If you are not used to adopt this approach, you need to do it more often and try different strategies until one of them work.”

Scenario 3

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. She decided, in the last lesson, to assign some problems that assess whether her students had grasped the idea of checking if the values that they found for the variables could be considered as a solution to the problem. She observed that a few students were giving answers like the one below.

> Your aunt Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper with different vacant plots being sold, all of them with an area of 800m^2. However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?



$x(x+20) = 800$

$x^2 + 20x = 800$

$x^2 + 20x - 800 = 0$

$\Delta = b^2 - 4ac$

$\Delta = 20^2 - 4 . 1 . (-800)$

$\Delta = 400 + 3200$

$\Delta = 3600$

$\sqrt{\Delta} = \pm 60$

$x = \dfrac{-b \pm \sqrt{\Delta}}{2a}$

$x = \dfrac{-20 + 60}{2} \begin{cases} \rightarrow 20 \\ \rightarrow -40 \end{cases}$

$x = 20 \rightarrow x + 20 = 40 \rightarrow \boxed{20 \text{ and } 40}$

$x = -40 \rightarrow x + 20 = -40 + 20 = -20 \rightarrow \boxed{-40 \text{ and } -20}$

Q6. What would be the best feedback to give to these students?

(a) Congratulations! Your response is almost completely correct. However, you did not complete it by giving a final correct answer for the sides of the plot. You need to establish whether the values that you found can be considered as a solution to the problem.

(b) Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think all the quadratic solutions can be accepted?

(c) Congratulations! Your response is almost completely correct. Only the last part needs to be checked. If the problem asks for the values of the side of a plot, can you accept negative values?

(d) Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.

Mrs Brown then wrote:

> Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think all the quadratic solutions can be accepted?

Q7. What would be the best thing for Mrs Brown to do in the next lesson?

(a) Present a correct solution of this problem on the board to the whole class questioning the students to generate each step.

(b) Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments on their first attempt.

(c) Ask those students who did not solve this problem fully to tackle a similar problem, reminding them to take account of her written comments on their previous work.

(d) Ask one student who solved this problem correctly to present their solution on the board to the whole class.

Not so sure about what to do, Mrs Brown decided to ask those students who did not solve the problem fully to tackle a similar problem, reminding them to take account of her written comments on their previous work.
In the end of the lesson, she collected the problems and analysed them again very carefully. However, she observed that some students were still making the same mistake.

Talking to some teachers, they gave her different explanations of why this had happened.

Q8. In your opinion, which teacher gave the best explanation?

(a) Mr Styles: "That's normal. Some students have more difficulties than others and will need more time. With a little more practice they will get there."

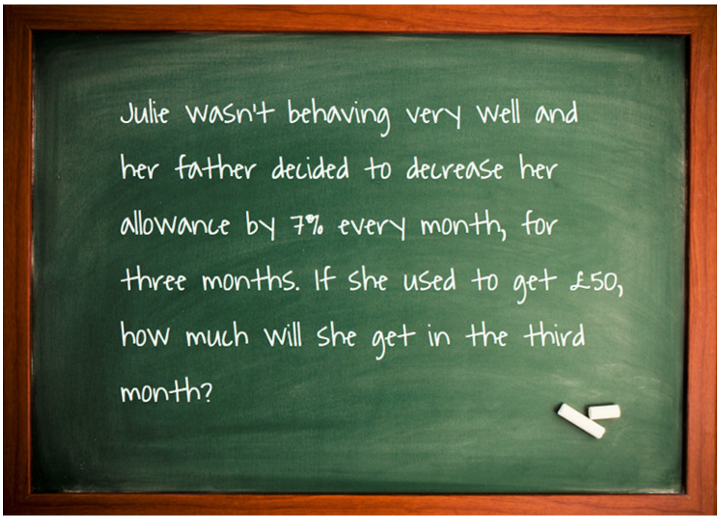
(b) Mrs Evans: "I believe that this happened because the feedback you gave was too long and complicated. Students don't like to read. Maybe you should give them a point to encourage them."

(c) Mrs Forbes: "In my opinion this happened because the students are not used to use teacher's written feedback. You need to help them to use feedback for their own learning."

(d) Mrs Johnson: "I think this happened because you jumped one step. You should have first given them the same problem and then a similar one to verify if they really learned."

Scenario 4

Mrs White was revising fractions and percentages with one of her 7th year classes. She started with some word problems involving percentages. She wrote on the board:



One of the students, started participating:

Anna: We need to find 7% of 50 and then multiply by 3.

Mrs White: And what do you need to do first?

Anna: Take it away from 50.

Mrs White records the two steps on the board saying:

"This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion.

"Instead of timesing by 0.07 loads of times, you could times by 0.21."

Q9. How could you best interpret Paul's answer?

(a) Paul is following the teacher suggestion and doing the process again and again.

(b) Paul is multiplying by 3 instead of using the power of 3.

(c) Paul has forgotten to take it away from 50.

(d) Paul is focusing on the 0.07 and the 3, without taking account of the situation.

Q10. In this case, what would be a good way to continue the discussion?

(a) "This may not work, Paul. But I will record your suggestion on the board so we can discuss it later"

(b) "0.21 is 0.07 times 3, right? Can you figure out whether this solution may or may not work in this problem, Paul?"

(c) "0.21, Paul? Tell me why this may work in this case. Can you show me how you came up with 0.21?"

(d) "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"

Scenario 5

Mr. Fitzgerald has been teaching his 7^{th} year students how to *divide by fractions*. In the next lesson, he decides to assess whether his students know how to divide by $\frac{1}{2}$. For that, he developed three problems:

I. Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

II. You have \$1.25 and may soon double your money. How much money will you end up with?

III. Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?

Before assigning the problems to his students, he showed you the problems and asked whether they are appropriate for his intention or not.

Q11. What do you think is the best advice to give to Mr. Fitzgerald?

(a) "I think you should review your assignment because some of your exercises are not assessing what you want them to assess."

(b) "Have you thought about including more questions because three are not sufficient for giving the information that you are looking for."

(c) "I think that you should review your assignment because none of your exercises are assessing what you want them to assess."

(d) "Why don't you include exercises which require students to divide by fractions other than $\frac{1}{2}$?"

Scenario 6

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 6 on her first quiz of the term and after a week or so, had a 7 in the second one.

Her father approached her because he was a little concerned as she usually gets more than that.

> Father: "Amy, I noticed that your marks in the last mathematics quizzes were not as high as they usually are. How can I help you, my darling?"
>
> Amy: "Don't worry, dad. Mr Hickman explained that there would soon be a test. The test will be divided into sections with each section representing a previous quiz. If I do better on a particular section than I did on the corresponding quiz, my grade on that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure about Mr Hickman's approach - why he would substitute the grades. He then decided to go to school to talk to him.

Q12. What do you think is the best explanation Mr Hickman could give to Amy's father?

(a) "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark along the term."

(b) "Although I give a mark to every quiz, they are designed for practice really. If I don't substitute the mark, students will not see a reason to study for the test"

(c) "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this way, the ultimate grade will represent learning."

(d) "The previous quizzes will help Amy guide her studying and will no doubt help her to learn the content by test time. Her mark will reflect what she really learned."

Mr Hickman then realises that other parents could be having the same doubt as Amy's father and decides to send them an email explaining his approach.
Some parents replied asking about how their children should use the past quizzes to improve.

Q13. How is the best way to advise these parents?

(a) "Students should review to the test by looking over their old quizzes to identify what they have struggled with in the past, and work more on those areas."

(b) "You should help your children to review for the test by going over their old quizzes and showing them how to solve the questions that they missed and asking them to practice those that they already got right at the first attempt."

(c) "Students need to learn how to solve all the exercises that they didn't get right in the first attempt. Therefore, I would advise them to go over all the past quizzes and re-do all exercises to make sure that they don't miss any."

(d) "Students should re-do all the quizzes many times to practice for the final test on the term. This way, they will for sure have a better mark."

B.2 Second version with 13 questions

This is the English version of the questionnaire that was presented to the teachers in PHASE 3A (section 6.2). Therefore, it already encompassed the suggestions made by the researchers during PHASE 2 (section 6.1).

Scenario 1

Mr. Smith was teaching his students how to solve inequalities. However, when he assigned an assessment activity, he observed that the majority of the class were making mistakes. See some examples of how they were solving them:

1. $17 - 2x > 25$

$-2x > 25 - 17$

$-2x > 8$

$x > -4$

2. $23 + x < 3x - 9$

$x - 3x < -9 - 23$

$-2x < -32$

$x < 16$

Q1. What is the best interpretation that can be made in relation to these patterns and the students' learning?

(a) These students understood the steps to solve inequalities, but had a small mistake in the end.

(b) These students solved as if they were equations, as they did not take the inequality into account.

(c) These students do not know how to solve inequalities, because they did not find the correct results in any of the exercises.

(d) These students did not pay attention to what the teacher explained, as they did not take the inequality into account.

Q2. Since the majority of students were having difficulties, what would be the best type of feedback to give to them?

(a) Give written comments to every student who did not solve it correctly, and let them try to solve it again in the next lesson.

(b) Go over that specific content to the whole group in the next lesson, and give the students another chance to re-do the exercises.

(c) Go over the content to the whole group in the next lesson, so they can be successful in the next assessment.

(d) Return the assessment to students and give them the opportunity to re-do the exercise in class, asking for the teacher's help when they deem necessary.

Mr Smith was not very sure about what to do, but he decided to go over the content to the whole group.

Q3. What would be the best way to make the aim clear to students when he starts the conversation in the next lesson?

(a) "In this lesson, we will go over the activity from last lesson because I noticed that some of you had trouble doing it."

(b) "Today, I will show you all how to solve inequalities because I noticed that some of you did it wrongly in the last lesson activity."

(c) "In our lesson today, we will correct the last lesson activity in the board because I noticed that some of you got it wrong."

(d) "We are going to start our lesson by solving some inequalities because I noticed that some students had had trouble with the activity from the last lesson."

Q4. Once the aim is clear, what would be the best thing for Mr Smith to do next?

(a) Repeat previous explanation and pass back the activity so students can try it again.

(b) Explain to students what kind of mistake they were making and pass back the activity so they can try it again.

(c) Pass back the activity to students commenting on the mistakes displayed and present the correct solution on the board.

(d) Call one student who got them right to come to the board to demonstrate and after that the others solve it in their notebook.

Scenario 2

Mrs. Andrews is used to giving written feedback to each student on almost all the test questions that she assigns. In the last test, she decided to change strategies, because she wanted to give her students the opportunity to discuss their ideas in pairs and try to solve the questions together.
When she returned the test to her students, she said:

> "This is the test I gave you in the last lesson. You will see that when you got a question right, I indicated it. When you didn't get it right, I did not write anything."
>
> "What I want you to do is: choose a friend, sit together and solve each question that you didn't get right. At the end of the lesson, I will collect the test so I can have a look again."

Working together was new to the students, and Mrs. Andrews observed that, instead of discussing the questions, students were copying the correct answers from their colleagues.
Later, she told the other teachers what had happened so they could give her some advice. Although she had managed to work around the unexpected situation, she still wanted to try this strategy again, at another opportunity.

Q5. What would be the best advice to give to Mrs. Andrews?

(a) "The best way to do this is by putting together those students with similar performance. This way, it is more likely that they will discuss and try to solve the questions together, instead of copying them from a colleague."

(b) "I believe that you need to remind them about this experience and explain that they did not do what you asked them. Make clear that you will try again, but in the next time they will not have another opportunity to re-do their test."

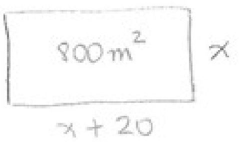
(c) "The most important part is to make clear to students what your intention with the activity is. That you are giving them a second opportunity of learning, and that you are putting them in pairs so they can discuss and help each other."

(d) "Working in pairs is not easy. You need to give some time for your students to learn and get used to it. If you are not used to put them in pairs, you need to do it more often and try different strategies until one of them work."

Scenario 3

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. In the last lesson, she decided to assign some problems that assess whether her students had grasped the idea of checking if the values that they found for the variables could be considered as a solution to the problem. She observed that a few students were giving answers like the one below.

> Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper with different vacant plots being sold, all of them with an area of 800m^2. However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?



$$x(x+20) = 800$$
$$x^2 + 20x = 800$$
$$x^2 + 20x - 800 = 0$$
$$\Delta = b^2 - 4ac$$
$$\Delta = 20^2 - 4 . 1 . (-800)$$
$$\Delta = 400 + 3200$$
$$\Delta = 3600$$
$$\sqrt{\Delta} = \pm 60$$

$$x = \frac{-b \pm \sqrt{\Delta}}{2a}$$
$$x = \frac{-20 \pm 60}{2} \begin{cases} \to 20 \\ \to -40 \end{cases}$$

$x = 20 \to x + 20 = 40 \to \boxed{20 \text{ and } 40}$

$x = -40 \to x + 20 = -40 + 20 = -20 \to \boxed{-40 \text{ and } -20}$

Q6. Based on this situation, what would be the best feedback to give to these students?

(a) Congratulations! Your response is almost correct. However, the sides of the plot that you presented are not correct. You need to analyse whether the values that you found can be considered as a solution to the problem.

(b) Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think all the quadratic solutions can be accepted?

(c) Congratulations! Your response is almost correct. You just need to check the last part. If the problem asks for the values of the side of a plot, can you accept negative values?

(d) Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.

Mrs Brown then wrote:

> Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think all the quadratic solutions can be accepted?

Q7. What would be the best thing for Mrs Brown to do in the next lesson, after giving written feedback to these students?

(a) Correct the problem on the board to the whole class asking students to generate each step.

(b) Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments.

(c) Ask those students who did not solve this problem fully to tackle a similar problem, asking them to take account of her written comments.

(d) Ask one student who solved this problem correctly to present their solution on the board to the whole class.

Not so sure about what to do, Mrs Brown decided to ask those students who did not solve the problem fully to tackle a similar problem, asking them to take account of her written comments.
In the end of the lesson, she collected the problems and analysed them again. However, she observed that some students were still making the same mistake.
Talking to other teachers, they gave her different explanations of why this had happened.

Q8. In your opinion, which teacher gave the best explanation?

(a) Mr Styles: "That's normal. Some students have more difficulties than others and will need more time. With a little more practice they will get there."

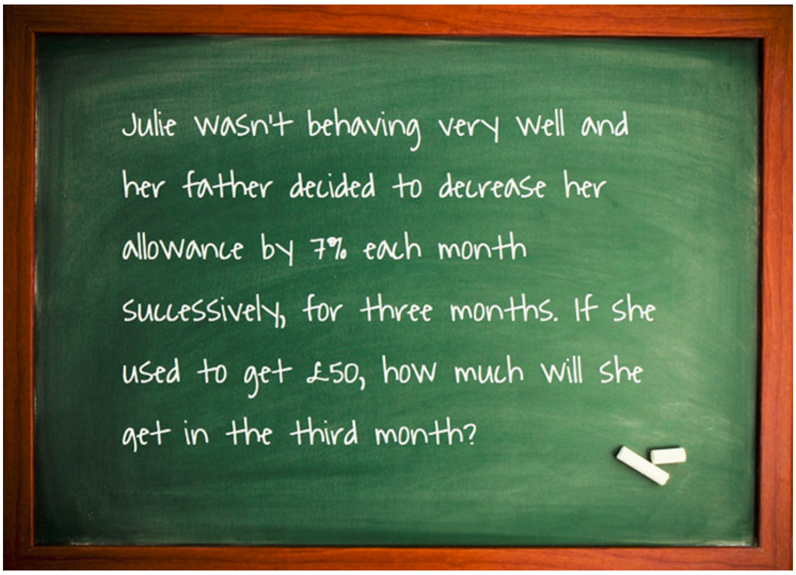
(b) Mrs Evans: "I believe that this happened because the feedback you gave was too long and complicated. Maybe you should have given some mark to encourage them."

(c) Miss Forbes: "Students are not used to use teacher's written feedback. You need to help them to use the information for their own learning."

(d) Mrs Johnson: "I think you jumped one step. You should have first given them the same problem and then a similar one to verify if they really learned."

Scenario 4

Mrs White was revising fractions and percentages with one of her 7th year classes. She started with some word problems involving percentages. She wrote on the board:



One of the students, started participating:

Anna: We need to find 7% of 50 and then multiply by 3.

Mrs White: And what do you need to do first?

Anna: Take it away from 50.

Mrs White records the two steps on the board saying:

"This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion.

Instead of timesing by 0.07 loads of times, you could times by 0.21.

Q9. How could you best interpret Paul's answer?

(a) Paul is following the teacher suggestion and doing the process again and again.

(b) Paul is multiplying by 3 instead of using the power of 3.

(c) Paul has forgotten to take it away from 50.

(d) Paul is focusing on the 0.07 and the 3, without taking account of the situation.

Q10. In this case, what would be a good way to continue the discussion?

(a) "0.21 may not work, Paul. But I will record your suggestion on the board so we can discuss it later"

(b) "0.21 is 3 times 0.07, right? Can you figure out whether this solution may or may not work in this problem, Paul?"

(c) "Can you show me how you came up with 0.21, Paul? Tell me why this may work in this case."

(d) "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"

Scenario 5

Mr Fitzgerald has been teaching his 7^{th} year students how to *divide by fractions*. In the next lesson, he decides to assess whether his students know how to divide by $\frac{1}{2}$. For that, he developed three problems:

I. Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

II. You have \$1.25 and may soon double your money. How much money will you end up with?

III. Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?

Before assigning the problems to his students, he showed you the problems and asked whether they are appropriate for his intention or not.

Q11. What do you think is the best advice to give to Mr. Fitzgerald?

(a) "I think you should review your assignment because some of your problems are not assessing what you want to assess."

(b) "Have you thought about including more questions because three are not sufficient for giving the information that you are looking for?"

(c) "I think you should review your assignment because none of your problems are assessing what you want to assess."

(d) "Why don't you include exercises which require students to divide by fractions other than $\frac{1}{2}$?"

Scenario 6

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 6 on her first quiz of the term and 7 in the second one.

Her father approached her because he was a little concerned as she usually gets more than that.

> Father: "Amy, I noticed that your marks in the last mathematics quizzes were not as high as they usually are. How can I help you, my darling?"
>
> Amy: "Don't worry, dad. Mr Hickman explained that there would soon be a test. The test will be divided into sections with each section representing a previous quiz. If I do better on a particular section than I did on the corresponding quiz, my grade on that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure about Mr Hickman's approach - why he would substitute the grades. He then decided to go to school to talk to him.

Q12. What do you think is the best explanation Mr Hickman could give to Amy's father?

(a) "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark along the term."

(b) "Although I give a mark to every quiz, they are designed for practice really. If I don't substitute the mark, students will not see a reason to study for the test"

(c) "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this way, the ultimate grade will represent what they have learnt in the end."

(d) "The previous quizzes will help Amy guide her studying and will help her, without a doubt to learn the content by test time. Her mark will reflect what she really learned."

Q13. And how do you think is the best way to advise students to use the tests to improve their learning?

(a) "You must review to the test by looking over your old quizzes to identify what they have struggled with in the past, and work more on those areas."

(b) "You must look for help to review for the test by going over your old quizzes and learning how to solve the questions that you missed and also practice those that you already got right at the first attempt."

(c) "You need to learn how to solve all the exercises that you didn't get right in the first attempt. Therefore, I would advise you to go over all the past quizzes and re-do all exercises to make sure that you don't miss any."

(d) "You must re-do all the quizzes many times to practice for the final test on the term. This way, you will for sure have a better mark."

B.3 First version with 24 questions

Scenario 1

Mr. Smith was teaching his students how to solve inequalities. However, when he assigned an assessment activity, he observed that the majority of the class were making mistakes. See some examples of how they were solving them:

1. $17 - 2x > 25$

$-2x > 25 - 17$

$-2x > 8$

$x > -4$

2. $23 + x < 3x - 9$

$x - 3x < -9 - 23$

$-2x < -32$

$x < 16$

Q1. What is the best interpretation we can make in relation to these patterns and the students' learning?

(a) These students understood the steps to solve inequalities, but had a small mistake in the end.

(b) These students solved as if they were equations, as they did not consider the inequality.

(c) These students do not know how to solve inequalities, because they did not find the correct results in any of the exercises.

(d) These students did not pay attention to what the teacher explained, as they did not consider the inequality.

Q2. Since the majority of students were having difficulties, what would be the best type of feedback to give to them?

(a) Give written comments to every student who did not solve it correctly, and let them try to solve it again in the next lesson.

(b) Re-explain that specific content to the whole group in the next lesson, and give the students another chance to re-do the exercises.

(c) Go over the content to the whole group in the next lesson, so they can be successful in the next assessment activity.

(d) Return the assessment to students and give them the opportunity to re-do the exercise in class, asking for the teacher's help when they deem necessary.

Mr Smith judged that the best strategy would be to go over the content to the whole group.

Q3. What would be the best way to make the aim clear to students when he starts the conversation in the next lesson?

(a) "In this lesson, we will go over the activity from last lesson because I noticed that many of you had trouble doing it."

(b) "Today, I will show you all how to solve inequalities because I noticed that many of you did it wrongly in the last lesson activity."

(c) "In our lesson today, we will correct the last lesson activity in the board because I noticed that the majority of you got it wrong."

(d) "We will begin our lesson remembering how to solve inequalities because I noticed that some students had had trouble solving them in the last class activity."

Q4. Once the aim is clear, what would be the best thing for Mr Smith to do next?

(a) Repeat previous explanation and pass back the activity so students can try it again.

(b) Explain to students what kind of mistake they were making and pass back the activity so they can try it again.

(c) Pass back the activity to students and present the correct solution on the board commenting on the mistakes displayed.

(d) Call one student who got them right to come to the board to demonstrate, and after that the others solve it in their notebook.

Scenario 2

Before moving forward with the content in her 9th grade groups, Mrs Andrews has decided to use a peer-assessment strategy to assess whether her students know how to identify and apply the definitions and properties of radius and/or diameter to calculate length and area of a circle.

For that, she prepared the following activity:

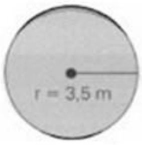
Circle and circumference activity

Name: ______________________________________.

Grade: 9th ______ **Date:** ___/___/_____.

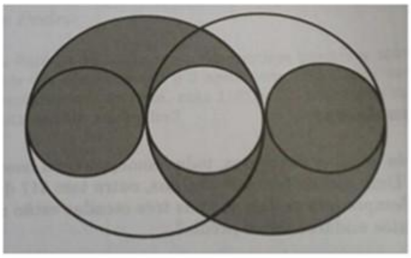
1. A disc has a diameter of 11.8 cm. The length of the circumference is approximately:
   a) 3.6 cm
   b) 37.1 cm
   c) 74.1 cm
   d) 11.8 cm

2. A pool has a circular shape. Knowing that its radius is equal to 3.5 m, it can be said that the area of the bottom of the pool is:
   a) 3.8465 m^2
   b) 38.465 m^2
   c) 384.65 m^2
   d) 3846.5 m^2



Use $\pi = 3.14$

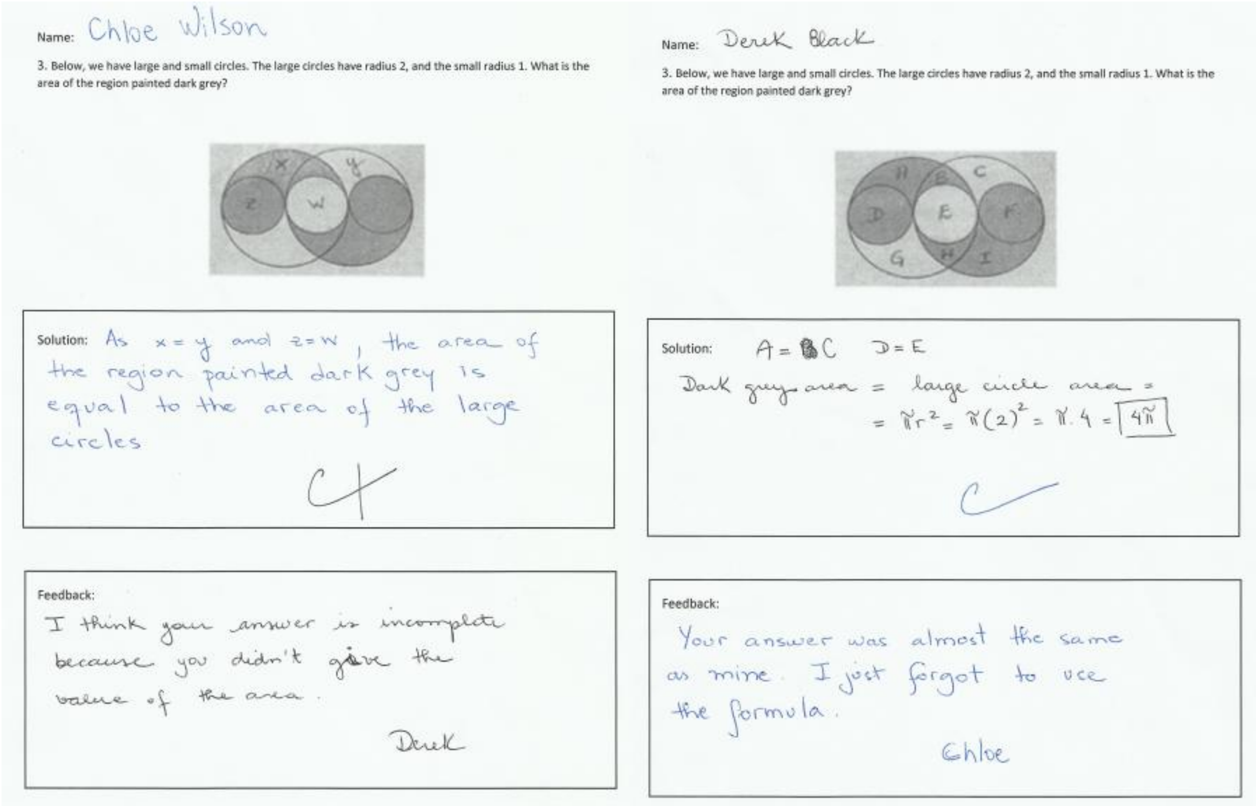
3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?



At the beginning of the lesson, Mrs Andrews handed the activity for the students and asked them to solve individually. When they finished, she said:

> "Now, you will exchange the activity with a colleague. You will analyse their answers, decide whether they are correct, incomplete or wrong and write a comment explaining your rationale. After that, I will collect the activities so I can analyse them myself."

When she was analysing, Mrs Andrews read interesting comments, like Chloe and Derek's:



Name: Chloe Wilson

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

Solution: As $x = y$ and $z = w$, the area of the region painted dark grey is equal to the area of the large circles

C+

Feedback: I think your answer is incomplete because you didn't give the value of the area.

Derek

Name: Derek Black

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

Solution: $A = C$ $D = E$

Dark grey area = large circle area =

$= \pi r^2 = \pi(2)^2 = \pi . 4 = \boxed{4\pi}$

C

Feedback: Your answer was almost the same as mine. I just forgot to uce the formula.

Chloe

Q5. Analysing the solution and comment, which is the best way Mrs Andrews can interpret Chloe's learning?

(a) Chloe knew how to calculate the area, but for some reason did not used the formula.

(b) Although Chloe did not present the calculation, her reasoning was correct.

(c) Chloe did not know how to calculate the area and only commented about the formula because she saw Derek's solution.

(d) Even though her first answer was incomplete, Chloe learned when she had the opportunity to assess Derek.

In the next lesson, Mrs Andrews brought all the activities back. Before passing it back to the students, she mentioned:

> "Yesterday, I had the opportunity to analyse all activities from our last lesson and I found very interesting solutions and comments..."

Q6. Which would be the best way to help students use their colleague's information in order to improve learning?

(a) "Now, I will pass them back so you re-do them using your colleague's comments. If you do not understand or do not agree with something, discuss it with your colleague."

(b) "Now, I will pass them back so you can re-do them analysing your colleague's comments. If you have any questions, ask me."

(c) "Now, I will pass them back for you can share your ideas and compare your solution strategies."

(d) "Now, I will pass them back for you to analyse your colleague's comments in order to have a better result in the next assessment."

When she concluded this step, Mrs Andrews observed that students have focussed their discussions around question 3 only. She then decided to talk about it with her colleagues so they could together, analyse the activity as a whole. Below, you see Mr Richards and Miss Lee's opinions.

Mr Richards: "I think that question 3 generates more discussions because it is more open and allows different solving strategies."

Miss Lee: "I, in turn, think it is because of questions 1 and 2, because they are multiple-choice. The student will select an option and that is it. That is why I do not give my students multiple-choice questions."

Mr Richards: "But questions 1 and 2 were in accordance with Mrs Andrews's intention - to assess whether her students know how to use radius and diameter to calculate length and area."

Miss Lee: "I do not agree because in question 2, actually, it is only assessing whether the student knows the position of the comma when they multiply decimals."

Q7. Which is the best way of analysing their contributions?

(a) Mr Richards presented a relevant analysis because he approached both the structure of question 3 and Mrs Andrews's goals.

(b) Miss Lee presented an important analysis because she highlighted the problem with the format of questions 1 and 2, which explains the attention to question 3.

(c) Both presented a significant analysis, as they both included important elements in their speeches.

(d) Both presented an incomplete analysis, which was insufficient to explain the problem presented.

At the end of the process, Mrs Andrews felt that she should verify whether the students had understood which the learning intentions were, having the peer-assessment activity as reference.

Q8. Which would be the best instruction Mrs Andrews could give?

(a) "Can you tell me what content we have been working in the last lessons? I will give you the last lesson activity so you can review the exercises and the feedbacks and try to identify which were our learning intentions."

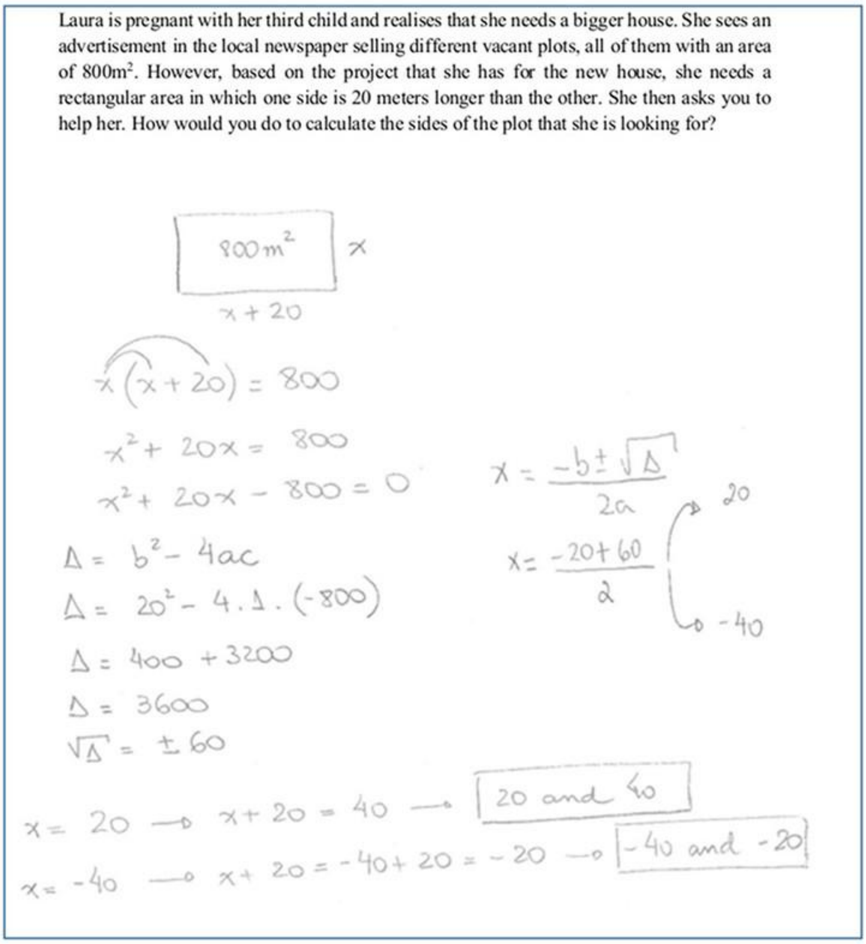
(b) "Do you remember what we have been learning about circumference? I will return the last lesson activity so you can have a look at the exercises and feedbacks and identify what we learned."

(c) "Do you remember the learning intentions from these few last lessons? I will write them in the board and pass back the last lesson activity so you can identify them in the exercises and feedbacks."

(d) "Can you tell me what we have been learning about circumference? I will ask you to use last lesson activity and the feedbacks in order to remember. After that, we will write your ideas on the board."

Scenario 3

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. She decided, in the last lesson, to assign some problems to assess whether her students had grasped the idea of checking if they could consider the values that they found for the variables as a solution to the problem. She observed that a few students were giving answers like the one below.


Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper selling different vacant plots, all of them with an area of 800m². However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?

$800 m^2$; x ; $x + 20$

$x(x+20) = 800$

$x^2 + 20x = 800$

$x^2 + 20x - 800 = 0$

$\Delta = b^2 - 4ac$

$\Delta = 20^2 - 4.1.(-800)$

$\Delta = 400 + 3200$

$\Delta = 3600$

$\sqrt{\Delta} = \pm 60$

$x = \frac{-b \pm \sqrt{\Delta}}{2a}$

$x = \frac{-20 \pm 60}{2}$ → 20 ; → -40

$x = 20 \rightarrow x + 20 = 40 \rightarrow \boxed{20 \text{ and } 40}$

$x = -40 \rightarrow x + 20 = -40 + 20 = -20 \rightarrow \boxed{-40 \text{ and } -20}$

Q9. Based on this situation, what would be the best feedback to give to these students?

(a) Congratulations! Your response is almost correct. However, the sides of the plot that you presented are not correct. You need to analyse whether you can consider the values that you found as a solution to the problem.

(b) Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?

(c) Congratulations! Your response is almost correct. You just need to check the last part. If the problem asks for the values of the side of a plot, can you accept the negative values?

(d) Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.

Mrs Brown then wrote:

> "Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think all the values you found can be accepted?"

Q10. What would be the best thing for Mrs Brown to do in the next lesson, after giving the written feedback to these students?

(a) Correct the problem on the board to the whole class asking students to generate each step.

(b) Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments.

(c) Ask those students who did not solve this problem fully to tackle a similar problem, asking them to take account of her written comments.

(d) Ask one student who solved this problem correctly to present their solution on the board to the whole class.

Mrs Brown considered that the best strategy would be to ask those students who did not solve the problem fully to tackle a similar problem, asking them to take account of her written comments.
In the end of the lesson, she collected the problems and analysed them again. However, she observed that some students were still making the same mistake.
Talking to other teachers, they gave her different explanations of why this had happened.

Q11. In your opinion, which teacher gave the best explanation?

(a) Mr Styles: “That’s normal. Some students have more difficulties than others do and will need more time. With a little more practice they will get there.”

(b) Mrs Evans: “I believe that this happened because the feedback you gave was too long and complicated. Maybe you should have given some mark to encourage them.”

(c) Miss Forbes: “Students are not used to use teacher’s written feedback. You need to help them to use the information for their own learning.”

(d) Mrs Johnson: “I think you jumped one step. You should have first given them the same problem and then a similar one to verify if they really learned.”

Mr Ledster arrived afterwards and wanted to participate in the conversation. He said:

> “In my opinion, this happened because the students are not used to using the teacher’s feedback. It is one of the teacher’s role to help students use assessment information to improve their learning.”

The other teachers found his comment very interesting and wanted to know how they could do that.

Q12. How would be the best way to answer these teachers?

(a) “A good way of doing that is to give written feedback to your students with a certain frequency. When they get used to it, they will end up developing this ability.”

(b) “There is not a good way of teaching this to students. Either they know how to interpret the feedback or they do not. That is why some students did the exercise correctly in the second attempt and others did not.”

(c) “A good way of doing that would be the teacher show the students how he/she would do that. The teacher would read the feedback with them and show how they could interpret it step by step.”

(d) “A good way of doing that would be to encourage students to give feedback to each other first. Having the opportunity to comment on a colleague’s work helps on developing the ability to interpret the teacher’s feedback.”

Scenario 4

Mrs White wanted to revise fractions and percentages with one of her 9th year classes. For that, she decided to ask students to develop some problems so they could solve them together in the classroom. However, for students be able to do the activity it was necessary to establish some criteria that the questions should meet.

Q13. How would be the best way to define these criteria?

(a) Devise the criteria together with the students.

(b) Define the criteria and explain them in detail to the students.

(c) Present some criteria and come up with a final version with the students.

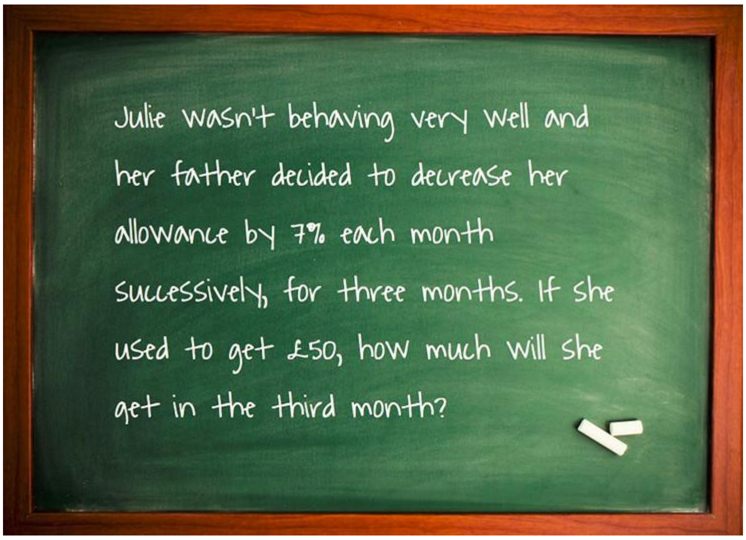
(d) Bring the criteria and define some value for each of them together with the students.

Together with students, they defined the following criteria:

The students should develop, in pairs, 4 questions meeting the following criteria:

1. The questions should approach the content of fractions and/or percentages;
2. The questions should be contextualised;
3. The question can be either open-ended or multiple-choice;
4. You should present at least one solution for each question.

In the next lesson, she decided to select some of those questions to solve with the students on the board. She started with the following:



One of the students started participating:

Anna: We need to find 7% of 50 and then multiply by 3.

Mrs White: And what do you need to do first?

Anna: Take it away from 50.

Mrs White records the two steps on the board saying:

> "This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion.

> "Instead of timesing by 0.07 loads of times, you could times by 0.21."

Q14. How could we best interpret Paul's answer?

(a) Paul is following the teacher suggestion and doing the process again and again.

(b) Paul is focusing on the 0.07 and the 3, without taking account of the situation.

(c) Paul is multiplying by 3 instead of using the power of 3.

(d) Paul has forgotten to take it away from 50.

Q15. In this case, what would be the best feedback to give to Paul?

(a) "0.21 may not work, Paul. But I will record your suggestion on the board so we can discuss it later."

(b) "0.21 is 3 times 0.07, right? Can you figure out whether this solution may or may not work in this problem, Paul?"

(c) "Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case."

(d) "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"

And the dialogue continued:

Mrs White: Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case.

Paul: The problem says three months. 3 times 0.07 is 0.21.

Mrs White: Yes, I agree with you: 0.21 is 3 times 0.07, but do you think this is what the problem is asking?

Paul: Well... the problem says 3 months, that's why I thought I had to multiply by 3.

Mrs White: It says 3 months, Paulo. But it's 7% in each month, and not all in the end of the third month, ok?

Paul: Ah... I get it. So we need first to take away 7% and then 7% again and 7% again.

Q16. In this case, how does the teacher should continue the discussion in order to verify if Paul learned with the feedback given?

(a) "Has everybody understood? Let's do it together in the board?"

(b) "So let's do this: you solve it in your notebook and when I show it on the board, you check if you got the correct answer."

(c) "Ok. You take 7% of 50. Then, from what is left, you take 7% again and so on. Do you want to solve it on the board so everyone can see it?"

(d) "Let's do this: you calculate 21% and then you do like that, taking away 7, 7 and 7. When you finish, we can talk about the results."

Scenario 5

Mr Fitzgerald has been teaching his 7th year students how to divide by fractions. In the next lesson, he decides to assess whether his students know how to divide by $\frac{1}{2}$. For that, he developed three problems:

1. Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?
2. You have £1.25 and may soon double your money. How much will you end up with?
3. Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?

Before assigning the problems to his students, he showed you the problems and asked whether they are appropriate for his intention or not.

Q17. What do you think is the best advice to give to Mr. Fitzgerald?

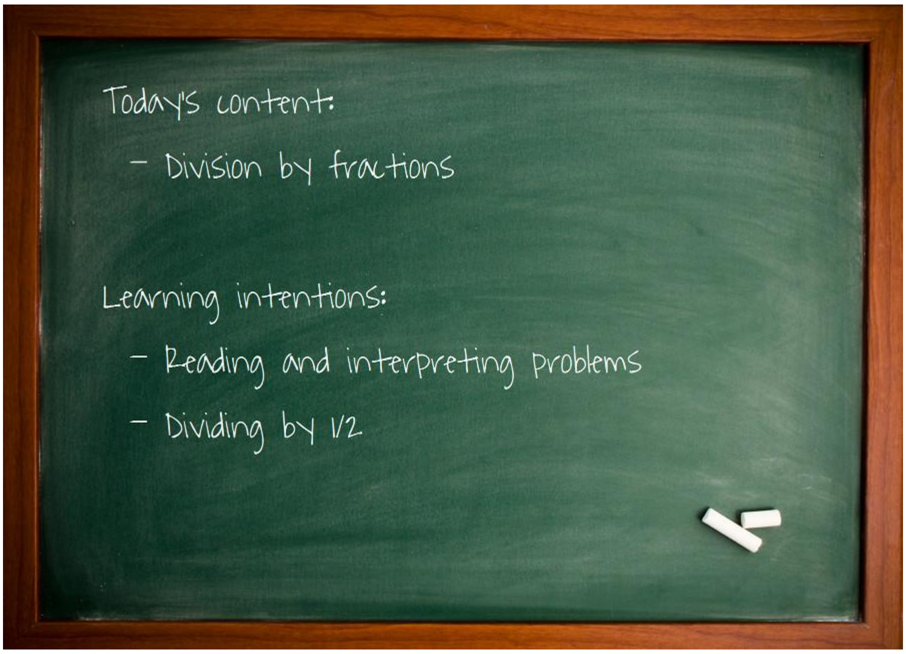
(a) "I think you should review your assignment because some of your problems are not assessing what you want to assess."

(b) "I think your activity is in accordance with your goal and therefore is ready to be given to students."

(c) "I think you should include more questions because three are not sufficient for giving the information that you are looking for."

(d) "I think you should include exercises which require students to divide by fractions other than $\frac{1}{2}$."

At the beginning of next lesson, Mr Fitzgerald felt it was important to make clear to the students which the learning intentions were and wrote on the board:



After that, the teacher gave the activity to students and put them into trios so they could solve the exercises together. Meanwhile, Mr Fitzgerald was walking around the room answering questions and observing students during the resolution of the activity. When passing by the trio Daniel, Carl and George, he heard the following conversation:

Daniel: Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

George: We just need to do $1 + \frac{1}{2}$ and divide by $\frac{1}{2}$.

Daniel: Yes.

Carl: No, guys. She wants to divide it between 2 friends. So, we have to do $1 + \frac{1}{2}$ and divide by 2.

Daniel: But it is written on the board that the exercises are to divide by half. So, we need to divide by half.

Carl: But it is also written that we need to read and interpret the problems, so I think it's a tricky one.

George: Ok... But even though... If she wants to divide into two friends, she wants to divide in half, which is what I said since the beginning.

Q18. Based on this conversation between the students, what can Mr Fitzgerald conclude about George's learning?

(a) George is dividing by half because is having trouble to interpret the problem.

(b) George is considering divide by half and divide in half as the same thing.

(c) George is dividing by half because of the learning intentions written on the board.

(d) As the problem did not address division by half, little can be said about George's learning.

The teacher then realised that some confusion was going on. However, he did not know if it was because they were considering divide by half and divide in half as the same thing or because of the learning intentions written on the board.

Q19. How could the teacher have avoided this confusion when dealing with the learning intentions?

(a) He should have used some strategy to check if the students had understood the intentions.

(b) He should have communicated the intentions orally so they would be clearer to students.

(c) He should have written only one intention on the board so it would be clearer to students.

(d) He should have written more details on the board to make the intentions clearer to students.

Facing this problem, Mr Fitzgerald decided to intervene in the conversation.

Mr Fitzgerald: How about you, Daniel? What do you think?

Daniel: I think we need to divide by half because to divide between two friends is to divide in half...

Mr Fitzgerald: So you agree with George... How about you, Carl? Do you agree with your colleagues?

Carl: No. I still think that if Claire wants to divide it between two friends, we need to divide it by 2.

Q20. How could Mr Fitzgerald proceed in order to help the group evaluate which way is the most appropriate?

(a) "I think we are at an impasse. You need to talk some more to reach a decision."

(b) "And why don't you do both calculations and then analyse the answers according to what the problem suggests?"

(c) "Let's read the problem again? Can we say that divide in half and divide by half is the same thing?"

(d) "Let's bring the discussion to the whole group and see if we can come to a conclusion?"

Scenario 6

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 5 on her first quiz of the term and 6 in the second one.
Her father approached her because he was a little concerned as she usually gets more than that.

> **Father:** "Amy, I noticed that your marks in the last maths quizzes were not as high as they usually are. How can I help you, my darling?"
>
> **Amy:** "Don't worry, dad. Mr Hickman explained that there would soon be a test. He will divide the test into sections with each section representing a previous quiz. If I do better on a particular section than I did on the corresponding quiz, my grade on that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure about Mr Hickman's approach - why he would substitute the grades. He then decided to go to school to talk to him.

Q21. What is the best explanation Mr Hickman could give to Amy's Father regarding his assessment procedure?

(a) "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark along the term."

(b) "Although I give a mark to every quiz, they are designed for practice really. If I don't substitute the mark, students will not see a reason to study for the test"

(c) "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this way, the ultimate grade will represent what they have learnt in the end."

(d) "The previous quizzes will help the students guide their studying and will help them to learn the content by test time. Their mark will better reflect what they really learned."

Q22. How is the best way to advise students to use the tests to improve their learning?

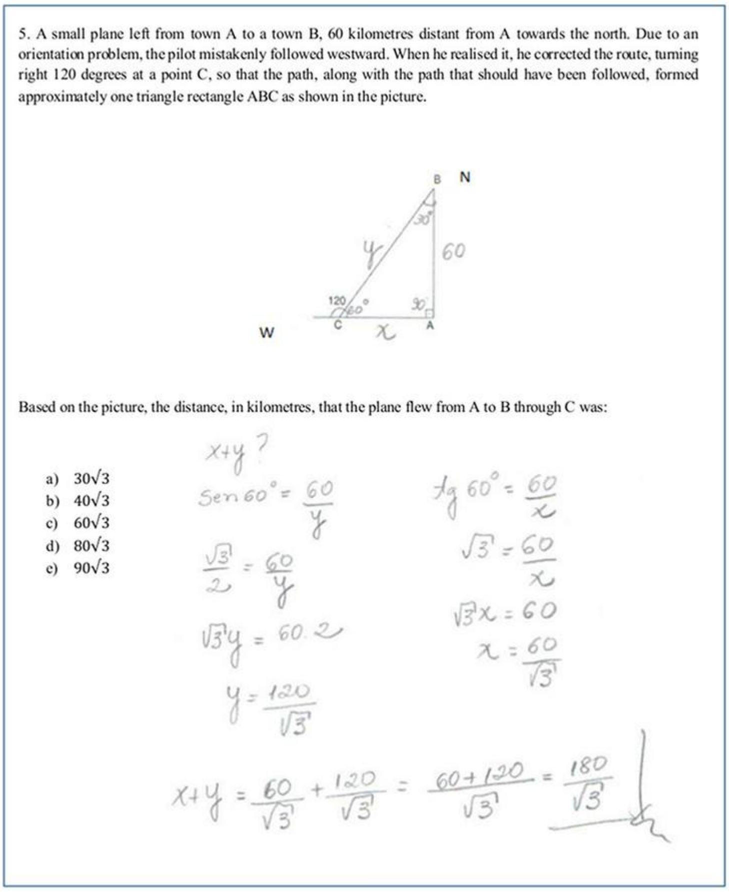
(a) "You should review to the test by looking over your old quizzes to identify what you have struggled with, and work more on those areas."

(b) "You should look for help to review for the test by going over your old quizzes to learn how to solve the questions that you missed and also re-do those that you already got right at the first attempt."

(c) "You should go over all the past quizzes and re-do all exercises to make sure that you don't miss any."

(d) "You should re-do all the quizzes many times to practice for the final test on the term. This way, you will for sure have a better mark."

After the test, Mr Hickman observed that Amy had overcome the difficulties presented in the first quiz but could still improve in some aspects of the content of the second quiz. Below, there is her solution of one of the questions about trigonometric identities:


5. A small plane left from town A to a town B, 60 kilometres distant from A towards the north. Due to an orientation problem, the pilot mistakenly followed westward. When he realised it, he corrected the route, turning right 120 degrees at a point C, so that the path, along with the path that should have been followed, formed approximately one triangle rectangle ABC as shown in the picture.

Based on the picture, the distance, in kilometres, that the plane flew from A to B through C was:

a) $30\sqrt{3}$
b) $40\sqrt{3}$
c) $60\sqrt{3}$
d) $80\sqrt{3}$
e) $90\sqrt{3}$

$x+y$?

$\text{Sen } 60° = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60.2$

$y = \frac{120}{\sqrt{3}}$

$tg\ 60° = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$

Facing this situation, Mr Hickman decided to write her a feedback in order to help her identify what was missing.

Q23. Which would be the best feedback to give Amy?

(a) Well done, Amy! Your reasoning is correct. You only missed the final part of the calculation. Now, you need to rationalise the denominator so you can find one of the answers provided.

(b) Well done, Amy! You understood the problem and correctly used the trigonometric identities. Can you think of what else you need to do then in order to find one of the answers provided?

(c) Well done, Amy! Your solution is correct. However, your final answer is not in the form I expected. Can you think of what else you need to do in order to find one of the answers provided?

(d) Well done, Amy! I can see that you understood that you needed to use the trigonometric identities to solve the problem. Now, you need to rationalise the denominator so you can find one of the answers provided.

Mr Hickman wrote the following feedback:

> Well done, Amy! You understood the problem and correctly used the trigonometric identities. Can you think of what else you need to do then in order to find one of the answers provided?

When returning the test, he called her in his table to talk to her.

Q24. What is the best thing for Mr Hickman to say when returning the commented test to Amy?

(a) "Amy, you will see that I wrote some comments in some of your questions because you still displayed some difficulties in some aspects. I suggest you to use these comments to try to solve these questions."

(b) "You have done well, Gabriela, because you did much better in the test. You see, that I wrote you some comments so you can learn even more."

(c) "Well done, Amy! You got a great mark in the test. Now I suggest you to read the comments so you can try to solve the incomplete questions."

(d) "Amy, you progressed a lot compared to the quizzes, but you can still improve in some aspects. I suggest you to read my comments to understand what is missing in some questions and try to complete them."

B.4 Field test version

In this section, I present the English version of the questionnaire that was used for the field test.

Scenario 1

Mr Smith was teaching his students how to solve inequalities. However, when he assigned an assessment activity, he observed that the majority of the class were making mistakes. See some examples of how they were solving them:

1. $17 - 2x > 25$

$$-2x > 25 - 17$$
$$-2x > 8$$
$$x > -4$$

2. $23 + x < 3x - 9$

$$x - 3x < -9 - 23$$
$$-2x < -32$$
$$x < 16$$

Q1. What is the best interpretation we can make in relation to these patterns and the students' learning?

(a) These students understood the steps to solve inequalities, but had a small mistake in the end.

(b) These students solved as if they were equations, as they did not consider the inequality.

(c) These students do not know how to solve inequalities, because they did not find the correct results in any of the exercises.

(d) These students did not pay attention to what the teacher explained, as they did not consider the inequality.

Q2. Since the majority of students were having difficulties, what would be the best type of feedback to give to them?

(a) Go over the content to the whole group in the next lesson, so they can be successful in the next assessment activity.

(b) Return the assessment to students and give them the opportunity to re-do the exercise in class, asking for the teacher's help when they deem necessary.

(c) Re-explain that specific content to the whole group in the next lesson, and give the students another chance to re-do the exercises.

(d) Give written comments to every student who did not solve it correctly, and let them try to solve it again in the next lesson.

Mr Smith judged that the best strategy would be to go over the content to the whole group.

Q3. What would be the best way to make the aim clear to students when he starts the conversation in the next lesson?

(a) "In our lesson today, we will correct the last lesson activity in the board because I noticed that the majority of you got it wrong."

(b) "Today, I will show you all how to solve inequalities because I noticed that many of you did it wrongly in the last lesson activity."

(c) "In this lesson, we will go over the activity from last lesson because I noticed that many of you had trouble doing it."

(d) "We will begin our lesson remembering how to solve inequalities because I noticed that some students had had trouble solving them in the last class activity."

Q4. Once the aim has been made clear, what would be the best thing for Mr Smith to do next?

(a) Explain to students what kind of mistake they were making and pass back the activity so they can try it again.

(b) Repeat previous explanation and pass back the activity so students can try it again.

(c) Pass back the activity to students and present the correct solution on the board commenting on the mistakes displayed.

(d) Call one student who got them right to come to the board to demonstrate, and after that the others solve it in their notebook.

Scenario 2

Mr Fitzgerald has been teaching his 7^{th} year students how to divide by fractions. In the next lesson, he decided to assess whether his students know how to divide by $\frac{1}{2}$. For that, he developed three problems:

I. Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

II. You have £1.25 and may soon double your money. How much will you end up with?

III. Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?

Before assigning the problems to his students, he showed you the problems and asked whether they are appropriate for his intention or not.

Q5. What do you think is the best advice to give to Mr. Fitzgerald?

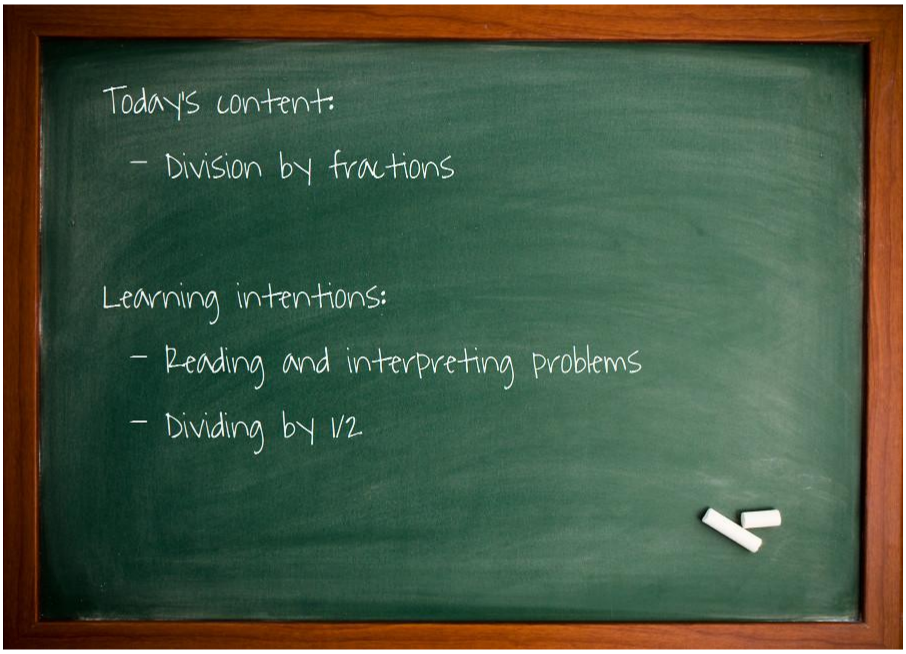
(a) "I think you should review your assignment because some of your problems are not assessing what you want to assess."

(b) "I think you should include exercises which require students to divide by fractions other than $\frac{1}{2}$."

(c) "I think your activity is in accordance with your goal and therefore is ready to be given to students."

(d) "I think you should include more questions because three are not sufficient for giving the information that you are looking for."

At the beginning of next lesson, Mr Fitzgerald felt it was important to make clear to the students which the learning intentions were and wrote on the board:



After that, the teacher gave the activity to students and put them into trios so they could solve the exercises together. Meanwhile, Mr Fitzgerald was walking around the room answering questions and observing students during the resolution of the activity. When passing by the trio Daniel, Carl and George, he heard the following conversation:

Daniel: "Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?"

George: "We just need to do $1 + \frac{1}{2}$ and divide by $\frac{1}{2}$."

Daniel: "Yes."

Carl: "No, guys. She wants to divide it between 2 friends. So, we have to do $1 + \frac{1}{2}$ and divide by 2."

Daniel: "But it is written on the board that the exercises are to divide by a half. So, we need to divide by a half."

Carl: "But it is also written that we need to read and interpret the problems, so I think it's a tricky one."

George: "Ok... But even though... If she wants to divide into two friends, she wants to divide by a half, which is what I said since the beginning."

Q6. Based on this conversation between the students, what can Mr Fitzgerald conclude about George's learning?

(a) George is dividing by a half because is having trouble to interpret the problem.

(b) George is considering divide by a half and divide in half as the same thing.

(c) George is dividing by a half because of the learning intentions written on the board.

(d) As the problem did not address division by half, little can be said about George's learning.

The teacher then realised that some confusion was going on. However, he did not know if it was because they were considering divide by half and divide in half as the same thing or because of the learning intentions written on the board.

Q7. How could the teacher have avoided this confusion when dealing with the learning intentions?

(a) He should have used some strategy to check if the students had understood the intentions.

(b) He should have written only one intention on the board so it would be clearer to students.

(c) He should have communicated the intentions orally so they would be clearer to students.

(d) He should have written more details on the board to make the intentions clearer to students.

Facing this problem, Mr Fitzgerald decided to intervene in the conversation.

Mr Fitzgerald: "How about you, Daniel? What do you think?"

Daniel: "I think we need to divide by half because to divide between two friends is to divide in half..."

Mr Fitzgerald: "So you agree with George... How about you, Carl? Do you agree with your colleagues?"

Carl: "No. I still think that if Claire wants to divide it between two friends, we need to divide it by 2."

Q8. How could Mr Fitzgerald proceed in order to help the group evaluate which way is the most appropriate?

(a) "I think you have an impasse here. You need to talk some more to reach a decision."

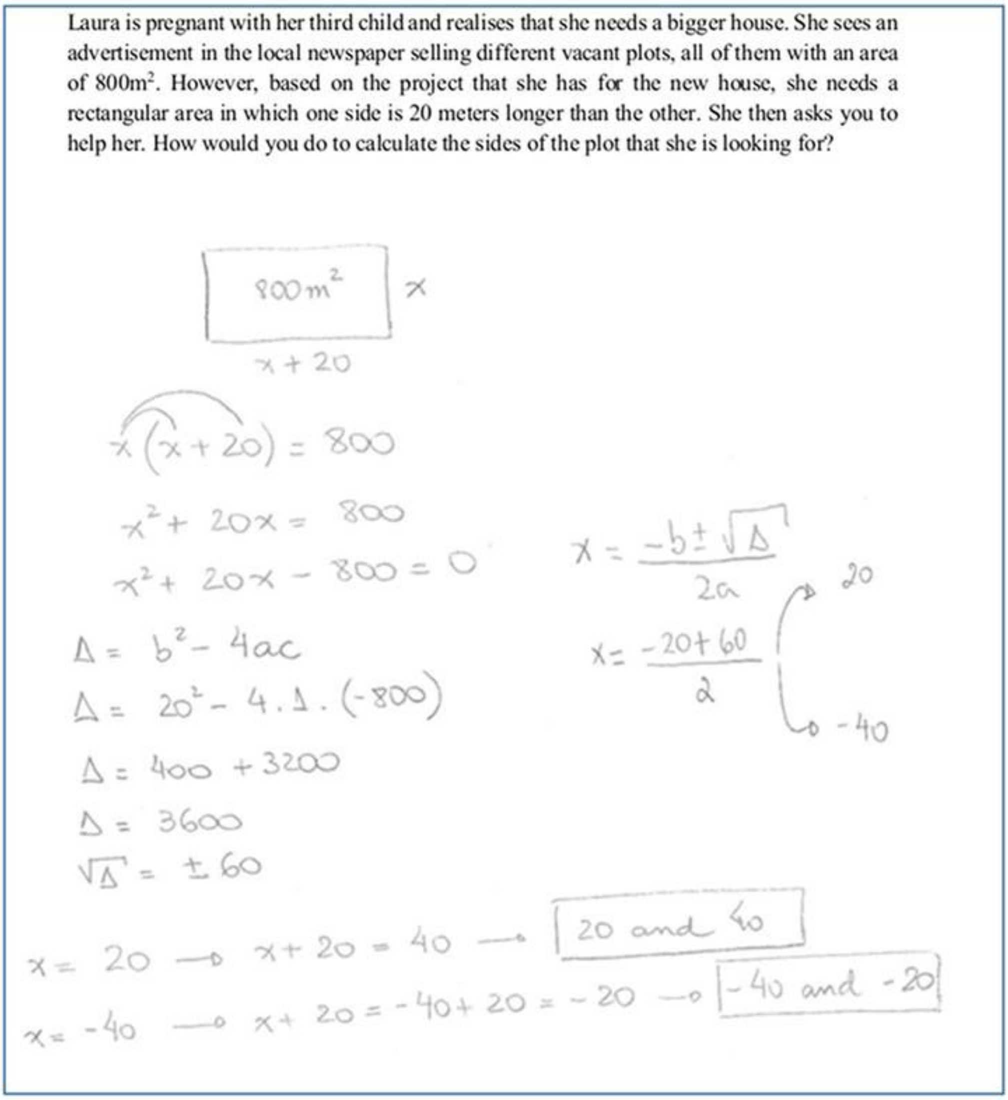
(b) "Let's read the problem again? Can we say that divide in half and divide by half is the same thing?"

(c) "And why don't you do both calculations and then analyse the answers according to what the problem suggests?"

(d) "Let's bring the discussion to the whole group and see if we can come to a conclusion?"

Scenario 3

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. She decided, in the last lesson, to assign some problems to assess whether her students had grasped the idea of checking if they could consider the values that they found for the variables as a solution to the problem. She observed that a few students were giving answers like the one below.


Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper selling different vacant plots, all of them with an area of 800m². However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?

$800m^2$, x, $x + 20$

$x(x+20) = 800$

$x^2 + 20x = 800$

$x^2 + 20x - 800 = 0$

$\Delta = b^2 - 4ac$

$\Delta = 20^2 - 4 . 1 . (-800)$

$\Delta = 400 + 3200$

$\Delta = 3600$

$\sqrt{\Delta} = \pm 60$

$x = \dfrac{-b \pm \sqrt{\Delta}}{2a}$

$x = \dfrac{-20 + 60}{2}$ → 20, → −40

$x = 20 \rightarrow x + 20 = 40 \rightarrow$ $\boxed{\text{20 and 40}}$

$x = -40 \rightarrow x + 20 = -40 + 20 = -20 \rightarrow$ $\boxed{\text{-40 and -20}}$

Q9. Based on this situation, what would be the best feedback to give to these students?

(a) Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?

(b) Congratulations! Your response is almost correct. However, the sides of the plot that you presented are not correct. You need to analyse whether you can consider the values that you found as a solution to the problem.

(c) Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.

(d) Congratulations! Your response is almost correct. You just need to check the last part. If the problem asks for the values of the side of a plot, can you accept the negative values?

Mrs Brown then wrote:

> Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?

Q10. Based on the feedback that Mrs Brown wrote, what would be the best thing for her to do in the next lesson?

(a) Ask one student who solved this problem correctly to present their solution on the board to the whole class.

(b) Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments.

(c) Correct the problem on the board to the whole class asking students to generate each step.

(d) Ask those students who did not solve this problem fully to tackle a similar problem, asking them to take account of her written comments.

Mrs Brown considered that the best strategy would be to ask those students who did not solve the problem fully to tackle a similar problem, asking them to take account of her written comments. In the end of the lesson, she collected the problems and analysed them again. However, she observed that some students were still making the same mistake.

Talking to other teachers, they gave her different explanations of why this had happened.

Q11. In your opinion, which teacher gave the best explanation?

(a) Mr Styles: “That’s normal. Some students have more difficulties than others do and will need more time. With a little more practice they will get there.”

(b) Mrs Evans: “I believe that this happened because the feedback you gave was too long and complicated. Maybe you should have given some mark to encourage them.”

(c) Miss Forbes: “The students could not identify their mistake through your comment. You should have made clear that the problem was to consider the negative values.”

(d) Mrs Johnson: “I think you jumped one step. You should have first given them the same problem and then a similar one to verify if they really learned.”

Mr Ledster arrived afterwards and wanted to participate in the conversation. He said:

> “In my opinion, this happened because the students are not used to using the teacher’s feedback. It is one of the teacher’s role to help students use assessment information to improve their learning.”

The other teachers found his comment very interesting and wanted to know how they could do that.

Q12. How would be the best way to answer these teachers?

(a) “There is not a good way of teaching this to students. It is better to give written feedback to your students with a certain frequency. When they get used to it, they will end up developing this ability.”

(b) “A good way of doing that would be to encourage students to give feedback to each other first. Having the opportunity to comment on a colleague’s work helps on developing the ability to interpret the teacher’s feedback.”

(c) “There is not a good way of teaching this to students. Either they know how to interpret the feedback or they do not. That is why some students did the exercise correctly in the second attempt and others did not.”

(d) “A good way of doing that would be the teacher show the students how he/she would do that. The teacher would read the feedback with them and show how they could interpret it step by step.”

Scenario 4

Before moving forward with the content in her 9th grade groups, Mrs Andrews has decided to use a peer-assessment strategy to assess whether her students know how to identify and apply the definitions and properties of radius and/or diameter to calculate length and area of a circle.

For that, she prepared the following activity:

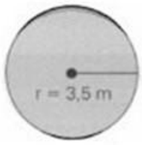
Circle and circumference activity

Name: ______________________________.

Grade: 9th ______ **Date:** ___/___/_____.

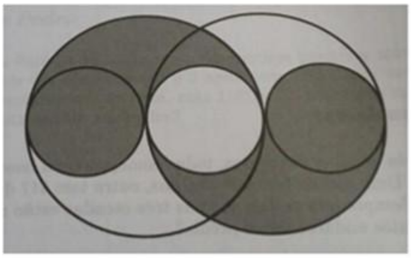
1. A disc has a diameter of 11.8 cm. The length of the circumference is approximately:
   a) 3.6 cm
   b) 37.1 cm
   c) 74.1 cm
   d) 11.8 cm

2. A pool has a circular shape. Knowing that its radius is equal to 3.5 m, it can be said that the area of the bottom of the pool is:
   a) 3.8465 m^2
   b) 38.465 m^2
   c) 384.65 m^2
   d) 3846.5 m^2



Use π = 3.14

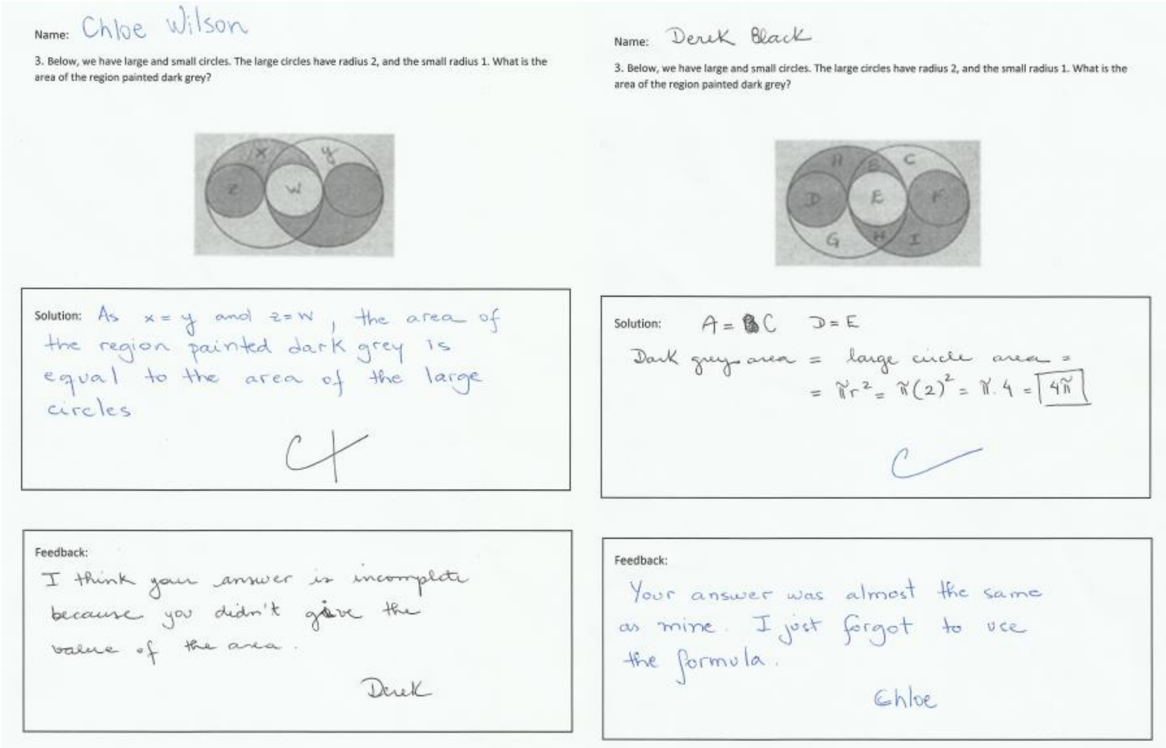
3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?



At the beginning of the lesson, Mrs Andrews handed the activity for the students and asked them to solve individually. When they finished, she said:

> "Now, you will exchange the activity with a colleague. You will analyse their answers, decide whether they are correct, incomplete or wrong and write a comment explaining your rationale. After that, I will collect the activities so I can analyse them myself."

When she was analysing, Mrs Andrews read interesting comments, like Chloe and Derek's:



Name: Chloe Wilson

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

Solution: As $x = y$ and $z = w$, the area of the region painted dark grey is equal to the area of the large circles

C+

Feedback: I think your answer is incomplete because you didn't give the value of the area.

Derek

Name: Derek Black

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

Solution: $A = C$ $D = E$

Dark grey area = large circle area =

$= \pi r^2 = \pi(2)^2 = \pi . 4 = \boxed{4\pi}$

C

Feedback: Your answer was almost the same as mine. I just forgot to use the formula.

Chloe

Q13. Analysing Chloe's solution and her comment on Derek's solution, which is the best way Mrs Andrews can interpret Chloe's learning?

(a) Even though her first answer was incomplete, Chloe learned when she had the opportunity to assess Derek.

(b) Chloe did not know how to calculate the area and only commented about the formula because she saw Derek's solution.

(c) Although Chloe did not present the calculation, her reasoning was correct.

(d) Chloe knew how to calculate the area, but for some reason did not use the formula.

In the next lesson, Mrs Andrews brought all the activities back. Before returning them to the students, she mentioned:

> "Yesterday, I had the opportunity to analyse all activities from our last lesson and I found very interesting solutions and comments..."

Q14. Which would be the best way to help students use their partner's information in order to improve learning?

(a) “Now, I will pass them back for you to analyse your partner’s comments in order to have a better result in the next assessment.”

(b) “Now, I will pass them back so you can re-do them using your partner’s comments. If you do not understand or do not agree with something, discuss it with your partner.”

(c) “Now, I will pass them back for you can share your ideas and compare your solution strategies”.

(d) “Now, I will pass them back so you can re-do them analysing your partner’s comments. If you have any questions, ask me”.

When she concluded this step, Mrs Andrews observed that students have focussed their discussions around question 3 only. She then decided to talk about it with her colleagues so they could together, analyse the activity as a whole. Below, you see Mr Richards and Miss Lee’s opinions.

Mr Richards: “I think that question 3 generates more discussions because it is more open and allows different solving strategies.”

Miss Lee: “I, in turn, think it is because of questions 1 and 2, because they are multiple-choice. The student will select an option and that is it. That is why I do not give my students multiple-choice questions.”

Mr Richards: “But questions 1 and 2 were in accordance with Mrs Andrews’s intention - to assess whether her students know how to use radius and diameter to calculate length and area.”

Miss Lee: “I do not agree because in question 2, actually, it is only assessing whether the student knows the position of the comma when they multiply decimals.”

Q15. Which is the best way of analysing their contributions?

(a) Mr Richards presented a relevant analysis because he approached both the structure of question 3 and Mrs Andrews’s goals.

(b) Miss Lee presented an important analysis because she highlighted the problem with the format of questions 1 and 2, which explains the attention to question 3.

(c) Both presented a significant analysis, as they both included important elements in their speeches.

(d) Both presented an incomplete analysis, which was insufficient to explain the problem presented.

At the end of the process, Mrs Andrews felt that she should verify whether the students had understood which the learning intentions were, using the peer-assessment activity as a reference point.

Q16. Which would be the best instruction Mrs Andrews could give to the students?

(a) "Do you remember what we have been learning about measuring circles? I will return the last lesson activity so you can have a look at the exercises and feedbacks and identify what we learned."

(b) "Do you remember the learning intentions from these few last lessons? I will write them in the board and pass back the last lesson activity so you can identify them in the exercises and feedbacks."

(c) "Can you tell me what we have been learning about measuring circles? I will ask you to use last lesson activity and the feedbacks in order to remember. After that, we will write the ideas on the board."

(d) "Can you tell me what content we have been working in the last lessons? I will give you the last lesson activity so you can review the exercises and the feedbacks and try to identify which were our learning intentions."

Scenario 5

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 5 on her first quiz of the term and 6 in the second one.
Her father approached her because he was a little concerned as she usually gets more than that.

Father: "Amy, I noticed that your marks in the last mathematics quizzes were not as high as they usually are. How can I help you, my darling?"

Amy: "Don't worry, dad. Mr Hickman explained that there would soon be a test. He will divide the test into sections with each section representing a previous quiz. If I do better on a particular section than I did on the corresponding quiz, my grade on that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure about Mr Hickman's approach – why he would substitute the grades. He then decided to go to school to talk to him.

Q17. What is the best explanation Mr Hickman could give to Amy's Father regarding his assessment procedure?

(a) "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark along the term."

(b) "Although I give a mark to every quiz, they are designed for practice really. If I don't substitute the mark, students will not see a reason to study for the test."

(c) "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this way, the ultimate grade will represent what they have learnt in the end."

(d) "The previous quizzes will help the students guide their studying and will help them to learn the content by test time. Their mark will better reflect what they really learned."

Q18. How is the best way to advise students to use the tests to improve their learning?

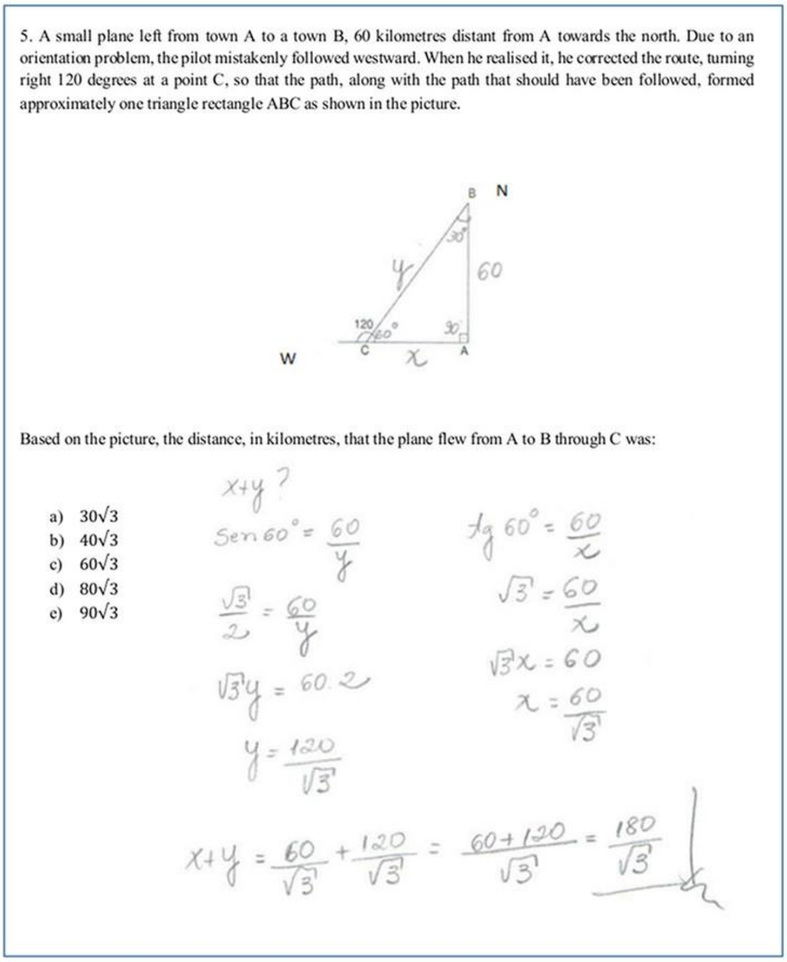
(a) “You should review to the test by looking over your old quizzes to identify what you have struggled with, and work more on those areas.”

(b) “You should look for help to review for the test by going over your old quizzes to learn how to solve the questions that you missed and also re-do those that you already got right at the first attempt.”

(c) “You should re-do all the quizzes many times to practice for the final test on the term. This way, you will for sure have a better mark.”

(d) “You should go over all the past quizzes and re-do all exercises to make sure that you don’t miss any.”

After the test, Mr Hickman observed that Amy had overcome the difficulties presented in the first quiz but could still improve in some aspects of the content of the second quiz. Below, there is her solution of one of the questions about trigonometric identities:



5. A small plane left from town A to a town B, 60 kilometres distant from A towards the north. Due to an orientation problem, the pilot mistakenly followed westward. When he realised it, he corrected the route, turning right 120 degrees at a point C, so that the path, along with the path that should have been followed, formed approximately one triangle rectangle ABC as shown in the picture.

Based on the picture, the distance, in kilometres, that the plane flew from A to B through C was:

a) $30\sqrt{3}$
b) $40\sqrt{3}$
c) $60\sqrt{3}$
d) $80\sqrt{3}$
e) $90\sqrt{3}$

$x+y$?

$\text{sen } 60° = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60.2$

$y = \frac{120}{\sqrt{3}}$

$\text{tg } 60° = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$

Facing this situation, Mr Hickman decided to write her a feedback in order to help her identify what was missing.

Q19. Which would be the best feedback to give Amy?

(a) Well done, Amy! Your reasoning is correct. Now, you need to rationalise the denominator so you can find one of the answers provided.

(b) Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?

(c) Well done, Amy! You understood that you needed to use the trigonometric identities to solve the problem. Now, you need to rationalise the denominator so you can find one of the answers provided.

(d) Well done, Amy! Your solution is almost correct. Can you think of what else you need to do in order to find one of the answers provided?

Mr Hickman wrote the following feedback:

> Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?

When returning the test, he called her in his table to talk to her.

Q20. What is the best thing for Mr Hickman to say when returning the commented test to Amy?

(a) "You have done well, Amy, because you did much better in the test. You will see that I wrote you some comments so you can learn even more."

(b) "Amy, you will see that I wrote some comments in some of your questions because you still displayed some difficulties in some aspects. I suggest you to use these comments to try to solve these questions."

(c) "Well done, Amy! You got a great mark in the test. Now I suggest you to read the comments so you can try to solve the questions that you missed."

(d) "Amy, you progressed a lot compared to the quizzes, but you can still improve in some aspects. I suggest you to read my comments to understand what is missing in some questions and try to complete them."

Scenario 6

Mrs White wanted to revise fractions and percentages with one of her 9th year classes. For that, she decided to ask students to develop some problems so they could solve them together in the classroom. However, for students be able to do the activity it was necessary to establish some criteria that the questions should meet.

Q21. How would be the best way to define these criteria?

(a) Devise the criteria together with the students.

(b) Define the criteria and explain them in detail to the students.

(c) Present some criteria and come up with a final version with the students.

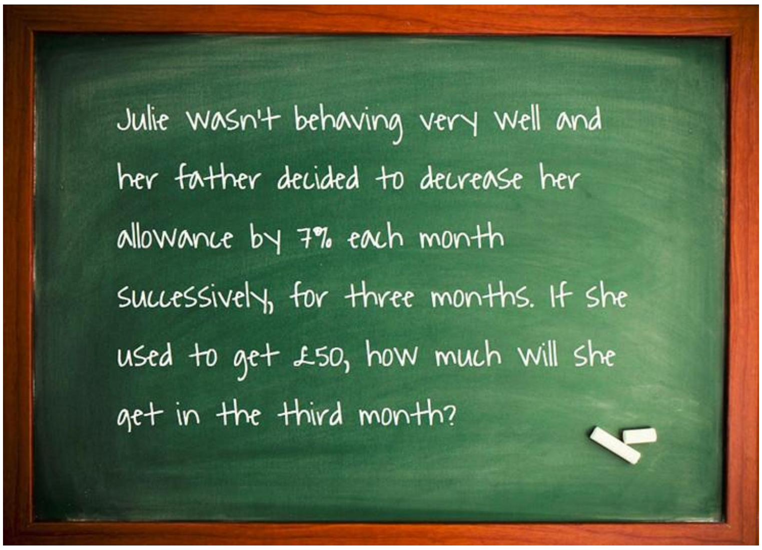
(d) Bring the criteria and define some weight for each of them together with the students.

Together with students, they defined the following criteria:

The students should develop, in pairs, 4 questions meeting the following criteria:

1. The questions should approach the content of fractions and/or percentages;
2. The questions should be contextualised;
3. The question can be either open-ended or multiple-choice;
4. You should present at least one solution for each question.

In the next lesson, she decided to select some of those questions to solve with the students on the board. She started with the following:



One of the students started participating:

Anna: "We need to find 7% of 50 and then multiply by 3."

Mrs White: "And what do you need to do first?"

Anna: "Take it away from 50."

Mrs White records the two steps on the board saying:

> "This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion.

Paul: "Instead of timesing by 0.07 loads of times, you could times by 0.21."

Q22. How could we best interpret Paul's answer?

(a) Paul is following the teacher suggestion and doing the process again and again.

(b) Paul is focusing on the 0.07 and the 3, without taking account of the situation.

(c) Paul is multiplying by 3 instead of using the power of 3.

(d) Paul has forgotten to take it away from 50.

Q23. In this case, what would be the best feedback to give to Paul?

(a) "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"

(b) "0.21 is 3 times 0.07, right? Can you figure out whether this solution may or may not work in this problem, Paul?"

(c) "Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case."

(d) "0.21 may not work, Paul. But I will record your suggestion on the board so we can discuss it later."

And the dialogue continued:

Mrs White: "Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case."

Paul: "The problem says three months. 3 times 0.07 is 0.21."

Mrs White: "Yes, I agree with you: 0.21 is 3 times 0.07, but do you think this is what the problem is asking?"

Paul: "Well... the problem says 3 months, that's why I thought I had to multiply by 3."

Mrs White: "It says 3 months, Paul, but it's 7% in each month, and not all in the end of the third month, ok?"

Paul: "Ah... I get it. So we need first to take away 7% and then 7% again and 7% again."

Q24. In this case, how does the teacher should continue the discussion in order to verify if Paul learned with the feedback given?

(a) "So let's do this: you solve it in your notebook and when I show it on the board, you check if you got the correct answer."

(b) "Has everybody understood what Paul just said? Let's do it together on the board."

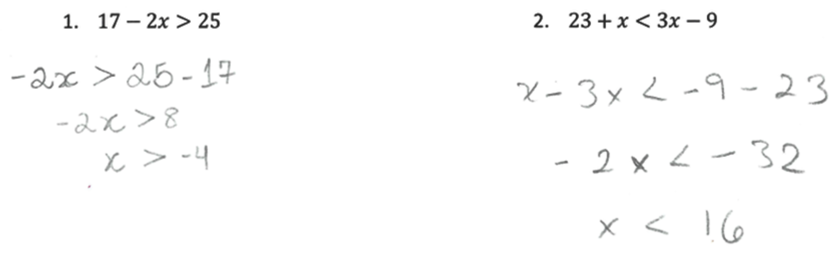
(c) "Yes. You take 7% of 50. Then, from what is left, you take 7% again and so on. Do you want to solve it on the board so everyone can see it?"

(d) "Let's do this: you calculate 21% and then you do like that, taking away 7, 7 and 7. When you finish, we can talk about the results."

B.5 Final version

Scenario 1

Mr Smith was teaching his 8th grade students how to solve inequalities. However, when he assigned an assessment activity, he observed that the majority of the class were making mistakes. See some examples of how they were solving them:


1. $17 - 2x > 25$

$-2x > 25 - 17$
$-2x > 8$
$x > -4$

2. $23 + x < 3x - 9$

$x - 3x < -9 - 23$
$-2x < -32$
$x < 16$

Q1. What is the best interpretation we can make in relation to these patterns and students' learning?

(a) These students understood the steps to solve inequalities, but made a small mistake at the end.

(b) These students solved the exercises as if they were equations, as they did not consider the inequality.

(c) These students do not know how to solve inequalities, because they did not find the correct result in any of the exercises.

(d) These students did not pay attention to what the teacher explained, as they did not consider the inequality.

Q2. Since the majority of students were having difficulties, what would be the best type of feedback to give to them?

(a) Go over the content whit the whole group in the next lesson, so they can be successful in the next assessment activity.

(b) Return the assessment to students and give them the opportunity to re-do the exercises in class, asking for the teacher's help when they deem necessary.

(c) Re-explain that specific content to the whole group in the next lesson, and give students another chance to re-do the exercises.

(d) Give written comments to every student who did not solve it correctly, and let them try to solve it again in the next lesson.

Mr Smith decided that the best strategy would be to go over the content with the whole group. At the beginning of the next lesson, he said:

> "We will begin our lesson remembering how to solve inequalities because I noticed that some students had trouble solving them in the last class activity."

Q3. Once the aim has been made clear, what would be the best thing for Mr Smith to do next?

(a) Explain to students what kind of mistake they were making and pass back the activity so they can try it again.

(b) Repeat the previous explanation and pass back the activity so students can try it again.

(c) Pass back the activity to students and present the correct solution on the board commenting on the mistakes displayed.

(d) Call one student who got the activity right to come to the board to demonstrate, and after that the others solve it in their notebook.

Scenario 2

Mr Fitzgerald has been teaching his 7th year students how to divide by fractions. For his next lesson, he has decided to assess whether his students know how to divide by $\frac{1}{2}$. For that, he has developed three problems:

1. Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?

2. You have £1.25 and may soon double your money. How much will you end up with?

3. Your mother is making cake and the recipe calls for two and a half cups of butter. How many sticks of butter will she need, knowing that each stick equals half a cup?

Before assigning them to his students, he decided to show you the problems and ask whether they are appropriate for his intention or not.

Q4. What do you think is the best advice to give to Mr Fitzgerald?

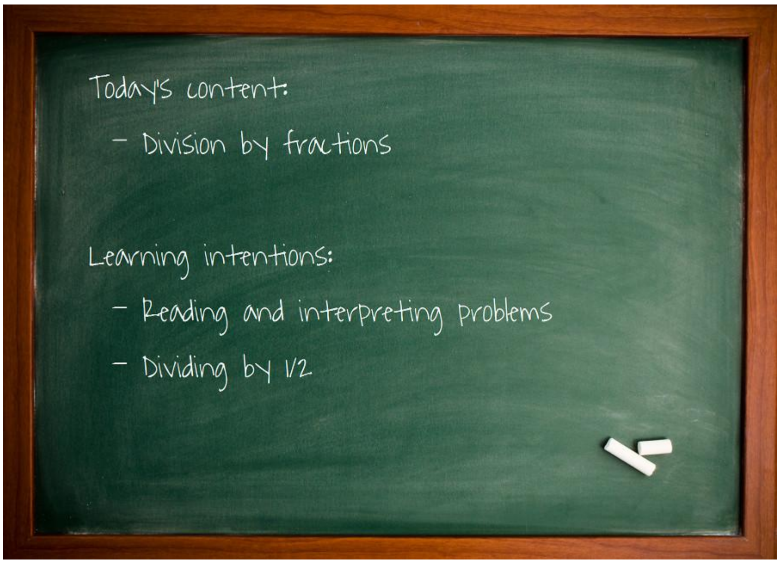
(a) "I think you should review your assignment because some of your problems are not assessing what you want to assess."

(b) "I think you should include exercises which require students to divide by fractions other than $\frac{1}{2}$."

(c) "I think your activity is in accordance with your goal and therefore is ready to be given to students."

(d) "I think you should include more questions because three are not sufficient for getting the information that you are looking for."

At the beginning of the next lesson, Mr Fitzgerald felt it was important to make clear to students what the learning intentions were and wrote on the board:



After that, Mr Fitzgerald gave the activity to students and put them into trios so they could solve the exercises together.
Meanwhile, Mr Fitzgerald walked around the room answering questions and observing students. When passing by the trio Daniel, Carl and George, he heard the following conversation:

Daniel: "Claire wants to split 1 pie and a half evenly between two friends. How much will each friend get?"

George: "We just need to do $1 + \frac{1}{2}$ and divide by $\frac{1}{2}$."

Daniel: "Yes."

Carl: "No, guys. She wants to divide it between 2 friends. So we have to do $1 + \frac{1}{2}$ and divide by 2."

Daniel: "But it is written on the board that the exercises are to divide by half. So we have to divide by half."

Carl: "But it is also written that we need to read and interpret the problems, so I think it's a tricky one."

George: "Ok... but even though... if she wants to divide into two friends, she wants to divide by half, which is what I have said since the beginning."

Q5. Based on this conversation between the students, what can Mr Fitzgerald conclude about George's learning?

(a) George is dividing by half because he is having trouble interpreting the problem.

(b) George is considering divide by half and divide in half as the same thing.

(c) George is dividing by half because of the learning intentions written on the board.

(d) As the problem did not address division by half, little can be said about George's learning.

Mr Fitzgerald then realised there was some confusion occurring.
However, he did not know if it was because they were considering divide by half and divide in half as the same thing or because of the learning intentions written on the board.

Q6. How could Mr Fitzgerald have avoided this confusion when dealing with the learning intentions?

(a) He should have used some strategy to check if students had understood the intentions.

(b) He should have written only one intention on the board to make it clearer for students.

(c) He should have communicated the intentions orally, instead of writing them, so they would be clearer for students.

(d) He should have written more details on the board to make the intentions clearer for students.

Facing this problem, Mr Fitzgerald decided to intervene in the conversation.

Mr Fitzgerald: "How about you, Daniel? What do you think?"

Daniel: "I think we need to divide by half because to divide between two friends is to divide in half..."

Mr Fitzgerald: "So you agree with George... how about you, Carl? Do you agree with your colleagues?"

Carl: "No. I still think that if Claire wants to divide it between two friends, we need to divide it by 2."

Q7. How could Mr Fitzgerald proceed in order to help the group evaluate which way is the most appropriate?

(a) "I think you have reached an impasse here. You need to talk some more to reach a decision."

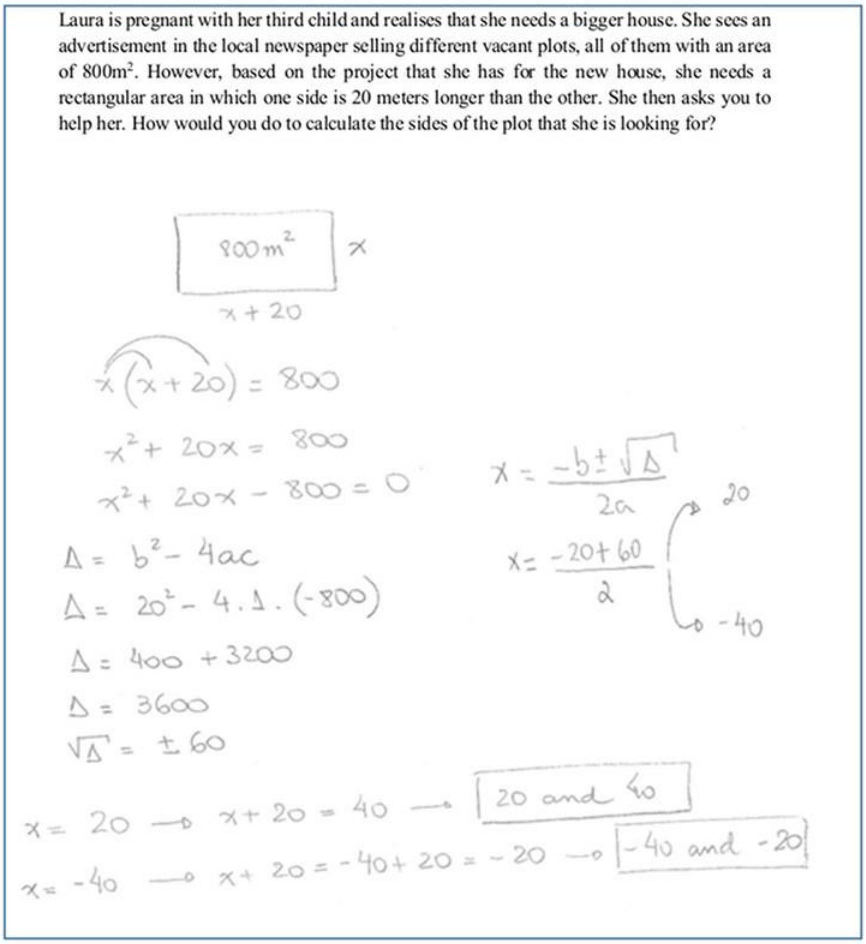
(b) "Let's read the problem again? Can we say that divide in half and divide by half are the same thing?"

(c) "Why don't you do both calculations and then analyse the answers according to what the problem suggests?"

(d) "Shall we bring the discussion to the whole group and see if we can come to a conclusion?"

Scenario 3

During the week, Mrs. Brown was teaching quadratic equations to her 9th year students. She decided, in the last lesson, to assign some problems to assess whether her students had grasped the idea of checking if they could consider the values that they found for the variables as a solution to the problem. She observed that a few students were giving answers like the one below.



Laura is pregnant with her third child and realises that she needs a bigger house. She sees an advertisement in the local newspaper selling different vacant plots, all of them with an area of $800m^2$. However, based on the project that she has for the new house, she needs a rectangular area in which one side is 20 meters longer than the other. She then asks you to help her. How would you do to calculate the sides of the plot that she is looking for?

$800m^2$ x

$x + 20$

$x(x+20) = 800$

$x^2 + 20x = 800$

$x^2 + 20x - 800 = 0$

$\Delta = b^2 - 4ac$

$\Delta = 20^2 - 4.1.(-800)$

$\Delta = 400 + 3200$

$\Delta = 3600$

$\sqrt{\Delta} = \pm 60$

$x = \frac{-b \pm \sqrt{\Delta}}{2a}$

$x = \frac{-20 \pm 60}{2}$ → 20, → -40

$x = 20 \rightarrow x + 20 = 40 \rightarrow \boxed{20 \text{ and } 40}$

$x = -40 \rightarrow x + 20 = -40 + 20 = -20 \rightarrow \boxed{-40 \text{ and } -20}$

Q8. Based on this situation, what would be the best feedback to give to these students?

(a) Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?

(b) Congratulations! Your response is almost correct. However, the sides of the plot that you presented are not correct. You need to analyse whether you can consider the values that you found as a solution to the problem.

(c) Congratulations! I can see that you figured out that you needed a quadratic equation to solve this problem. Now, you need to re-read the problem and analyse the answers you gave.

(d) Congratulations! Your response is almost correct. You just need to check the last part. If the problem asks for the values of the side of a plot, can you accept the negative values?

Mrs Brown then wrote:

> "Congratulations! You correctly solved the quadratic equation necessary to solve the problem. However, if the problem asks for the values of the sides of a plot, do you think you can accept all the values you found?"

Q9. Based on the feedback that Mrs Brown wrote, what would be the best thing for her to do in the next lesson?

(a) Ask one student who solved the problem correctly to present the solution on the board to the whole class.

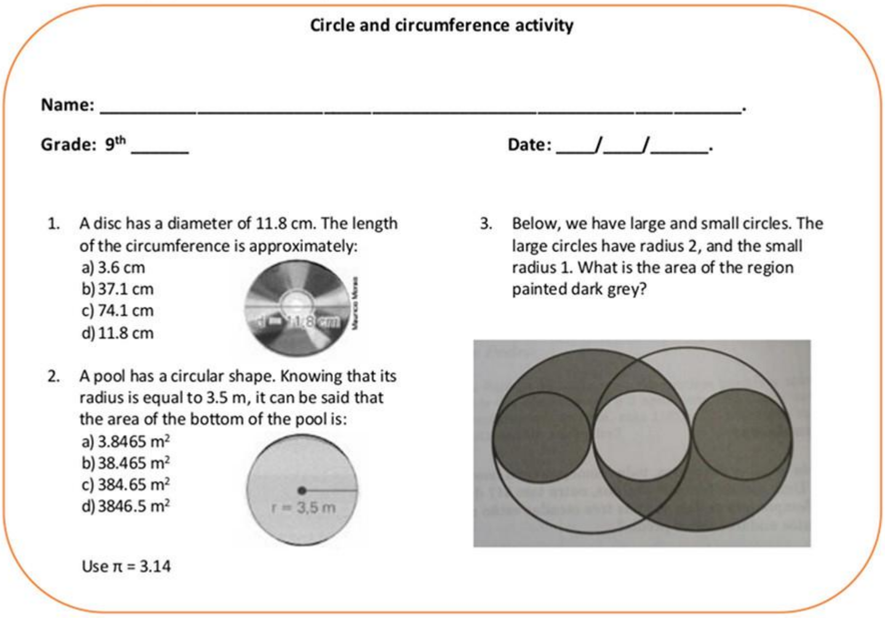
(b) Ask those students who did not solve this problem fully to tackle it again, paying attention to her written comments.

(c) Correct the problem on the board for the whole class, asking students to generate each step.

(d) Ask those students who did not solve this problem fully to tackle a similar problem, asking them to take account of her written comments.

Scenario 4

Before moving forward with the content in her 9th grade groups, Mrs Andrews has decided to use a peer-assessment strategy to assess whether her students know how to identify and apply the definitions and properties of radius and/or diameter to calculate length and area of a circle.
For that, she prepared the following activity:



Circle and circumference activity

Name: ____.

Grade: 9^{th} ______ **Date:** ___/___/_____.

1. A disc has a diameter of 11.8 cm. The length of the circumference is approximately:
   a) 3.6 cm
   b) 37.1 cm
   c) 74.1 cm
   d) 11.8 cm
2. A pool has a circular shape. Knowing that its radius is equal to 3.5 m, it can be said that the area of the bottom of the pool is:
   a) 3.8465 m^2
   b) 38.465 m^2
   c) 384.65 m^2
   d) 3846.5 m^2

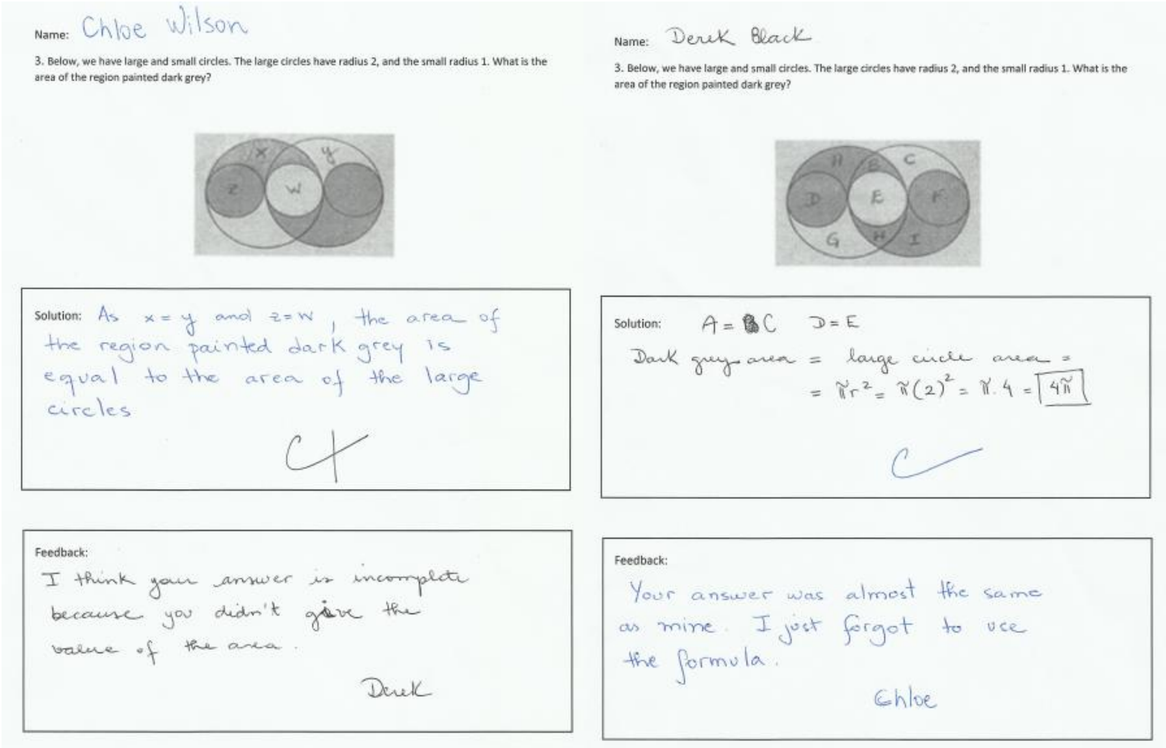
Use $\pi = 3.14$

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

At the beginning of the lesson, Mrs Andrews handed the activity to the students and asked them to solve individually. When they finished, she said:

> "Now, you will exchange the activity with a colleague. You will analyse their answers, decide whether they are correct, incomplete or wrong and write a comment explaining your rationale. After that, I will collect the activities so I can analyse them myself."

Whilst analysing, Mrs Andrews read interesting comments, like Chloe's and Derek's:


Name: Chloe Wilson

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

x y z w

Solution: As $x = y$ and $z = w$, the area of the region painted dark grey is equal to the area of the large circles

C+

Feedback: I think your answer is incomplete because you didn't give the value of the area.

Derek

Name: Derek Black

3. Below, we have large and small circles. The large circles have radius 2, and the small radius 1. What is the area of the region painted dark grey?

A B C D E F G H I

Solution: $A = C$ $D = E$

Dark grey area = large circle area =

$= \pi r^2 = \pi(2)^2 = \pi . 4 = \boxed{4\pi}$

C

Feedback: Your answer was almost the same as mine. I just forgot to uce the formula.

Chloe

Q10. Analysing Chloe's solution and her comment on Derek's solution, which is the best way Mrs Andrews can interpret Chloe's learning?

(a) Even though her first answer was incomplete, Chloe learned when she had the opportunity to assess Derek.

(b) Chloe did not know how to calculate the area and only commented about the formula because she saw Derek's solution.

(c) Although Chloe did not present the calculation, her reasoning was correct.

(d) Chloe knew how to calculate the area, but for some reason did not use the formula.

In the next lesson, Mrs Andrews brought all the activities back. Before returning them to the students, she said:

> "Yesterday, I had the opportunity to review all activities from our last lesson and I found very interesting solutions and comments. Now, I will pass them back so you can re-do them using your partner's comments. If you do

not understand or do not agree with something, please discuss it with your partner."

When she had concluded this step, Mrs Andrews observed that students focussed their discussions around question 3 only. She then decided to talk about it with her colleagues so that together they could analyse the activity as a whole. Below, you see Mr Richards' and Miss Lee's opinions.

Mr Richards: "I think that question 3 generated more discussion because it is more open and allows different solving strategies."

Miss Lee: "I, in turn, think it is because of questions 1 and 2 and the fact that they are multiple-choice. The student will select an option and that is it. That is why I do not give my students multiple-choice questions."

Mr Richards: "But questions 1 and 2 were in accordance with Mrs Andrews's intention - to assess whether her students know how to use radius and diameter to calculate length and area."

Miss Lee: "I do not agree because in question 2, actually, it is only assessing whether the student knows the position of the decimal point when multiplying decimals."

Q11. Which is the best way of analysing their contributions?

(a) Mr Richards presented a relevant analysis because he approached both the structure of question 3 and Mrs Andrews's goals.

(b) Miss Lee presented an important analysis because she highlighted the problem with the format of questions 1 and 2, which explains the attention to question 3.

(c) Both presented a significant analysis, as they both included important elements in their speeches.

(d) Both presented an incomplete analysis, which was insufficient to explain the problem presented.

At the end of the process, Mrs Andrews felt that she should verify whether students had understood what the learning intentions were, using the peer-assessment activity as a reference point.

Q12. Which would be the best instruction Mrs Andrews could give to the students?

(a) "Do you remember what we have been learning about measuring circles? I will return the activity from the last lesson so you can have a look at the exercises and feedback and identify what we learned."

(b) "Do you remember the learning intentions from the last few lessons? I will write them on the board and pass back the acticity from the last lesson so you can identify them in the exercises and feedback."

(c) "Can you tell me what we have been learning about measuring circles? I will ask you to use the acticity from the last lesson and the feedback in order to remember. After that, we will write the ideas on the board."

(d) "Can you tell me what content we have been learning in the last lessons? I will give you the acticity from the last lesson so you can review the exercises and the feedback and try to identify what our learning intentions were."

Scenario 5

Amy is a student who typically does very well in mathematics. However, she has a tendency to start slowly because she is often somewhat overwhelmed by new concepts. Amy received a 5 on her first quiz of the term and 6 in the second one.
Her father approached her because he was a little concerned as she usually gets more than that.

Father: "Amy, I noticed that your marks in the last maths quizzes were not as high as they usually are. How can I help you, my darlin?"

Amy: "Don't worry, dad. Mr Hickman explained that soon we will have an end-of-term test. He will divide the test into sections with each section representing a previous quiz. If I do better in a particular section than I did in the corresponding quiz, my grade for that section will replace the grade from the quiz."

Her father was happy with her confidence but was not very sure why Mr Hickman would replace the grades. He then decided to go to school to talk to the teacher.

Q13. What is the best explanation Mr Hickman could give to Amy's father regarding his assessment procedure?

(a) "What I am doing is what we call continuous recovery. That is, we need to give students different opportunities to recover their mark throughout the term."

(b) "Although I give a mark for every quiz, they are designed for practice really. If I don't replace the mark, students will not see a reason to study for the test."

(c) "The mark of the test is the one that matters the most because the quizzes are just for practice. By doing this, the grade on the test is the one which will represent what they have indeed learned."

(d) "The previous quizzes will help students guide their studying and will help them to learn the content by test time. Their mark will better reflect what they really learned."

Q14. Which is the best way to advise students to use the tests to improve their learning?

(a) "You should revise for the test by looking over your old quizzes to identify what you have struggled with, and work more on those areas."

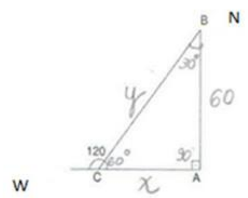
(b) "You should look for help to revise for the test by going over your old quizzes to learn how to solve the questions that you missed and also re-do those that you already got right at the first attempt."

(c) "You should re-do all the quizzes many times to practice for the final test of the term. This way, you will for sure have a better mark."

(d) "You should go over all the past quizzes and re-do all exercises to make sure that you don't miss any."

After the test, Mr Hickman observed that Amy had overcome the difficulties presented in the first quiz but could still improve in some aspects of the content of the second quiz. Below, there is her solution to one of the questions about trigonometric identities:

5. A small plane left from town A to a town B, 60 kilometres distant from A towards the north. Due to an orientation problem, the pilot mistakenly followed westward. When he realised it, he corrected the route, turning right 120 degrees at a point C, so that the path, along with the path that should have been followed, formed approximately one triangle rectangle ABC as shown in the picture.



Based on the picture, the distance, in kilometres, that the plane flew from A to B through C was:

a) $30\sqrt{3}$
b) $40\sqrt{3}$
c) $60\sqrt{3}$
d) $80\sqrt{3}$
e) $90\sqrt{3}$

$x+y$?

$\text{Sen}\, 60^\circ = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60 . 2$

$y = \frac{120}{\sqrt{3}}$

$\text{tg}\, 60^\circ = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$

Given the situation, Mr Hickman decided to write her feedback to help her identify what was missing.

Q15. Which would be the best feedback to give Amy?

(a) Well done, Amy! Your reasoning is correct. Now, you need to rationalise the denominator so you can find one of the answers provided.

(b) Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?

(c) Well done, Amy! You understood that you needed to use the trigonometric identities to solve the problem. Now, you need to rationalise the denominator so you can find one of the answers provided.

(d) Well done, Amy! Your solution is almost correct. Can you think of what else you need to do in order to find one of the answers provided?

Mr Hickman wrote the following feedback:

> "Well done, Amy! You correctly used the trigonometric identities to solve the problem. Can you think of what else you need to do in order to find one of the answers provided?"

When returning the test, he called her to his table to talk to her.

Q16. What is the best thing for Mr Hickman to say when returning the commented test to Amy?

(a) "You have done well, Amy, because you did much better in the test. You will see that I wrote you some comments so you can learn even more."

(b) "Amy, you will see that I wrote some comments in some of your questions because you still displayed some difficulties in some aspects. I suggest you to use these comments to try to solve these questions."

(c) "Well done, Amy! You got a great mark in the test. Now I suggest you to read the comments so you can try to solve the questions that you missed."

(d) "Amy, you progressed a lot compared to the quizzes, but you can still improve in some aspects. I suggest you to read my comments to understand what is missing in some questions and try to complete them."

Scenario 6

Mrs White wanted to revise fractions and percentages with one of her 9th year classes. For that, she decided to ask students to develop some problems so they could solve them together in the classroom.
However, for students to be able to do the activity it was necessary to establish some criteria that the questions should meet.

Q17. What would be the best way to define these criteria?

(a) Devise the criteria together with students.

(b) Define the criteria and explain them in detail to students.

(c) Present some criteria and come up with a final version with students.

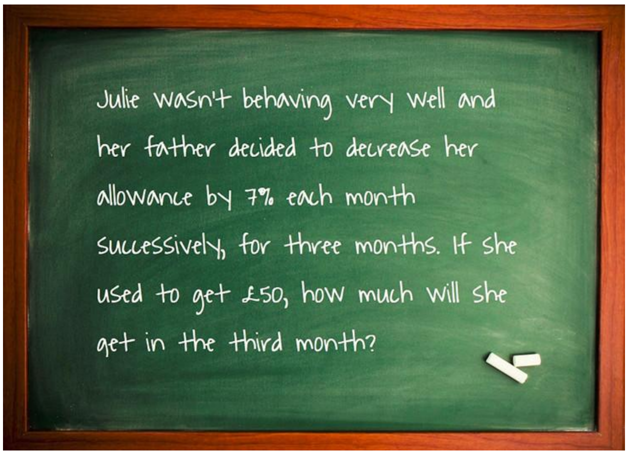
(d) Bring the criteria and define some weight for each of them together with the students.

Together with the students, they defined the following criteria:

The students should develop, in pairs, 4 questions meeting the following criteria:

1. The questions should approach the content of fractions and/or percentages;
2. The questions should be contextualised;
3. The question can be either open-ended or multiple-choice;
4. You should present at least one solution for each question.

In the next lesson, she decided to select some of those questions to solve with the students on the board.
She started with the following:



One student began participating:

Ana: "We need to find 7% of 50 and then multiply by 3."

Mrs White: "And what do you need to do first?"

Ana: "Take it away from 50."

Mrs White records the two steps on the board saying:

> "This gives the amount after the first month of decrease. We would need to do the same process again and again to get the amount after three months."

After that, Paul engaged in the discussion:

> "Instead of timesing by 0.07 loads of times, we could times by 0.21."

Q18. How could we best interpret Paul's answer?

(a) Paul is following the teacher suggestion and doing the process again and again.

(b) Paul is focusing on the 0.07 and the 3, without taking account of the situation.

(c) Paul is multiplying by 3 instead of using the power of 3.

(d) Paul has forgotten to take it away from 50.

Q19. In this case, what would be the best feedback to give to Paul?

(a) "If you multiply by 0.21, Paul, you will get 21% of 50 which is not what the problem is telling us. Can you think of another solution?"

(b) "0.21 is 3 times 0.07, right? Can you figure out whether this solution may or may not work in this problem, Paul?"

(c) "Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case."

(d) "0.21 may not work, Paul, but I will record your suggestion on the board so we can discuss it later."

And the dialogue continued:

Mrs White: “Can you show me how you came up with 0.21, Paul? Tell me why you think this may work in this case.”

Paul: “The problem says three months. 3 times 0.07 is 0.21.”

Mrs White: “Yes, I agree with you: 0.21 is 3 times 0.07, but do you think this is what the problem is asking?”

Paul: “Well... the problem says 3 months, that’s why I thought I had to multiply by 3.”

Mrs White: “It says 3 months, Paul, but it’s 7% in each month, and not all in the end of the third month, ok?”

Paul: “Ah ... I get it. So we need first to take away 7% and then 7% again and 7% again.”

Q20. In this case, how should the teacher continue the discussion in order to verify if Paul learned with the feedback given?

(a) “So let’s do this: you solve it in your notebook and when I show it on the board, you check if you got the correct answer.”

(b) “Has everybody understood what Paul just said? Let’s do it together on the board.”

(c) “Yes. You take 7% of 50. Then, from what is left, you take 7% again and so on. Do you want to solve it on the board so everyone can see it?”

(d) “Let’s do this: you calculate 21% and then you do like that, taking away 7, 7 and 7. When you finish, we can talk about the results.”

Answer key

Table 11.3: Answer key

Scenarios	Questions	Option A	Option B	Option C	Option D
1	1	3	4	1	2
	2	2	1	4	3
	3	3	4	2	1
2	4	4	3	1	2
	5	3	4	2	1
	6	4	2	1	3
	7	1	3	4	2
3	8	4	2	3	1
	9	1	4	2	3
4	10	4	1	3	2
	11	4	2	3	1
	12	3	1	4	2
5	13	2	1	3	4
	14	4	3	1	2
	15	1	4	2	3
	16	2	3	1	4
6	17	3	1	4	2
	18	3	4	1	2
	19	2	3	4	1
	20	1	2	4	3

C Questionnaire versions - Portuguese

C.1 Field test version

In this section, I present the Portuguese version of the questionnaire which was used for the field test.

Scenario 1

O professor João estava trabalhando inequações com seus alunos do 8ºano. Porém, quando ele passou uma atividade avaliativa, ele observou que a maioria dos alunos estava cometendo erros. Veja dois exemplos de como eles estavam resolvendo:

1. $17 - 2x > 25$

$-2x > 25 - 17$

$-2x > 8$

$x > -4$

2. $23 + x < 3x - 9$

$x - 3x < -9 - 23$

$-2x < -32$

$x < 16$

Q1. Qual é a melhor interpretação que pode ser feita em relação à aprendizagem dos alunos, a partir das respostas dadas?

(a) Esses alunos entenderam os passos para resolverem inequações, mas tiveram um erro no finalzinho.

(b) Esses alunos resolveram como se fossem equações, pois não levaram a desigualdade em consideração.

(c) Esses alunos não sabem resolver inequações, pois não obtiveram o resultado correto em nenhum dos dois exercícios.

(d) Esses alunos não prestaram atenção no que o professor ensinou sobre inequações, pois não levaram a desigualdade em consideração.

Q2. Uma vez que a maioria dos alunos apresentou dificuldade, qual seria a melhor forma de dar feedback para os alunos?

(a) Entregar a atividade para os alunos e dar a eles outra oportunidade para resolverem os exercícios, pedindo a ajuda do professor quando eles tiverem necessidade.

(b) Retomar o conteúdo para a turma toda na próxima aula para que eles possam se sair melhor na próxima avaliação.

(c) Reexplicar o conteúdo específico para a turma toda na próxima aula, e dar aos alunos uma outra chance de refazer os exercícios.

(d) Escrever um comentário para os alunos que não resolveram corretamente, e na próxima aula deixar que tentem resolver novamente.

O professor João decidiu que a melhor estratégia era retomar o conteúdo para a turma toda.

Q3. Qual seria a melhor forma de deixar seus objetivos claros para os alunos ao começar a próxima aula?

(a) "Na nossa aula de hoje, nós vamos corrigir a atividade da última aula no quadro porque eu observei que a maioria de vocês fez errado."

(b) "Hoje, eu vou mostrar para vocês como que se resolve inequações porque eu notei que muitos de vocês resolveram errado na atividade da última aula."

(c) "Nessa aula, nós vamos rever a atividade da aula passada porque eu observei que muitos de vocês tiveram dificuldade em resolvê-la."

(d) "Nós vamos começar a nossa aula relembrando como se resolve inequações porque eu notei que alguns alunos tiveram dificuldade de resolvê-las na atividade da aula passada."

Q4. Depois que o objetivo estiver claro, qual seria a melhor forma de abordar o conteúdo com os alunos?

(a) Explicar aos os alunos os erros que eles cometeram e devolver a atividade para que eles possam tentar de novo.

(b) Retomar o conteúdo da aula passada e devolver a atividade aos alunos para que eles possam tentar de novo.

(c) Devolver a atividade aos alunos e apresentar a solução correta no quadro comentando os erros que eles cometeram.

(d) Pedir a um aluno que resolveu corretamente para vir ao quadro e demonstrar ao grupo, e em seguida os outros alunos resolvem no próprio caderno.

Scenario 2

O professor André está trabalhando o assunto de divisão por frações com seus alunos do 7º ano. Ele decidiu que na próxima aula avaliará se eles sabem dividir por $\frac{1}{2}$. Para isso, ele elaborou alguns probleminhas:

I. Clara quer dividir uma torta e meia entre dois amigos. Quanto cada amigo vai receber?

II. Você tem R$1,25 e vai logo dobrar essa quantia. Com quanto você vai ficar no final?

III. Sua mãe está fazendo um bolo e a receita pede duas xícaras e $\frac{1}{2}$ de manteiga. Quantos tabletes de manteiga ela vai precisar sendo que cada tablete é equivalente a meia xícara?

Antes de passar os problemas aos seus alunos, ele resolveu te mostrar e perguntou se você acha os probleminhas apropriados para a intenção que ele tem.

Q5. O que você diria ao professor André?

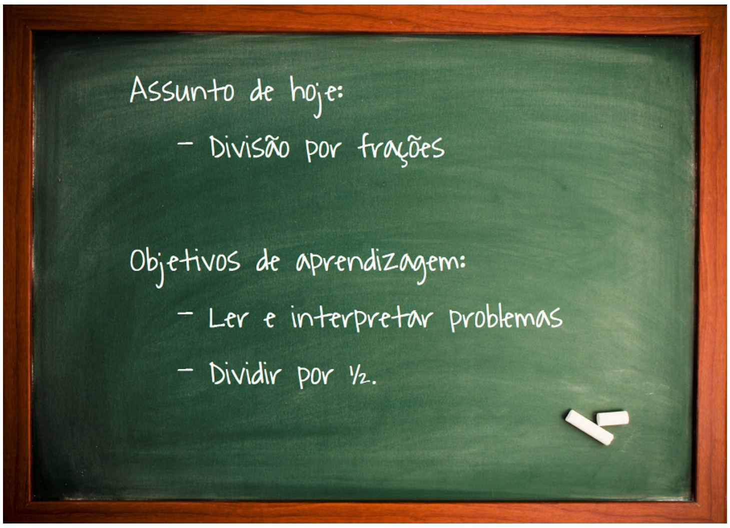
(a) "Eu acredito que você deva revisar sua atividade, pois alguns problemas não estão avaliando o que você quer avaliar."

(b) "Eu acredito que você deva incluir exercícios que exijam que os alunos dividam por outras frações além de $\frac{1}{2}$."

(c) "Eu acredito que a sua atividade esteja apropriada para o seu objetivo e, portanto, está pronta para ser aplicada."

(d) "Eu acredito que você deva incluir mais exercícios, pois três não são suficientes para obter a informação que você está querendo."

Ao começar a aula seguinte, o professor André achou importante deixar claro para os alunos quais eram os objetivos de aprendizagem daquela aula e escreveu no quadro:



Em seguida, ele entregou a atividade aos alunos e colocou-os em trios para que pudessem resolver juntos. Enquanto isso, o professor André ficou andando pela sala tirando dúvidas e observando os alunos durante a resolução da atividade. Ao passar pelo trio Daniel, Caio e Gustavo, ele ouviu a seguinte conversa:

Daniel: "Clara quer dividir uma torta e meia entre dois amigos. Quanto cada amigo vai receber?"

Gustavo: "É só a gente fazer $1 + \frac{1}{2}$ e dividir por $\frac{1}{2}$."

Daniel: "Isso."

Caio: "Não, gente. Ela quer dividir entre 2 amigos. Então tem que fazer $1 + \frac{1}{2}$ e dividir por 2."

Daniel: "Mas está lá escrito no quadro que os exercícios são para dividir por meio. Então a gente tem que dividir por meio."

Caio: "Mas também tem o outro objetivo que é ler e interpretar problemas, então eu acho que é uma pegadinha."

Gustavo: "Tá bom... Mas mesmo assim. Se ela quer dividir entre dois amigos, ela quer dividir na metade, isto é, dividir por meio, que é o que eu tinha dito desde o início."

Q6. Baseado na conversa entre os alunos, o que o professor André pode concluir sobre a aprendizagem do Gustavo?

(a) O Gustavo está dividindo por meio, pois está com dificuldade de interpretar o problema.

(b) O Gustavo está considerando dividir por meio e dividir no meio como sendo a mesma coisa.

(c) O Gustavo está dividindo por meio por causa dos objetivos de aprendizagem escritos no quadro.

(d) Como o problema não aborda divisão por meio, pouco se pode concluir sobre a aprendizagem do Gustavo.

O professor então percebeu que estava havendo uma confusão. Porém, ele ficou em dúvida se era porque os alunos estavam confundindo dividir por meio com dividir no meio, ou se era porque estavam sendo influenciados pelos objetivos escritos no quadro no início da aula.

Q7. Como o professor poderia ter evitado essa confusão ao lidar com os objetivos?

(a) Ele deveria ter utilizado alguma estratégia para verificar se os alunos haviam entendido os objetivos.

(b) Ele deveria ter escrito apenas um objetivo no quadro para que ficasse mais claro para os alunos entenderem.

(c) Ele deveria ter comunicado os objetivos oralmente para que eles ficassem mais claros para os alunos.

(d) Ele deveria ter escrito mais detalhes no quadro para que os objetivos ficassem mais claros para os alunos.

Diante dessa situação, o professor André decidiu intervir na conversa.

Prof. André: “E você, Daniel? O que você acha?”

Daniel: “Eu acho que a gente tem que dividir por meio porque dividir entre dois amigos é dividir no meio...”

Prof. André: “Então você concorda com o Gustavo... E você, Caio? Concorda com seus colegas?”

Caio: "Não. Eu ainda acho que se a Clara quer dividir entre dois amigos, nós temos que dividir por dois."

Q8. Como o professor André poderia dar prosseguimento de forma a ajudar o grupo a avaliar qual das formas é a mais apropriada?

(a) "Acho que vocês chegaram num impasse. Vocês precisam conversar um pouco mais para chegarem a uma decisão."

(b) "Vamos ler o problema novamente? Podemos dizer que dividir no meio e dividir por meio é a mesma coisa?"

(c) "E porque vocês não fazem as duas contas e analisam as respostas de acordo com o que o problema sugere?"

(d) "Vamos trazer a discussão para a turma toda e ver se conseguimos chegar a uma conclusão?"

Scenario 3

Durante a última semana, a professora Leila estava trabalhando equações de 2º grau com seus alunos do 9º ano. Na última aula, ela decidiu passar alguns problemas para avaliar se seus alunos haviam aprendido a verificar se as raízes da equação poderiam ser aceitas como solução do problema. Ela observou que alguns poucos alunos estavam respondendo como o exemplo abaixo.

Dona Eliana acabou de descobrir que está grávida do seu terceiro filho e decide construir uma casa maior. Ela ficou bem interessada em alguns terrenos de 800m² que encontrou nos classificados do jornal. Porém, baseada no projeto que ela tem para a nova casa, seria necessário um terreno retangular em que um dos lados fosse 20m maior do que o outro. Ela então resolve pedir sua ajuda. Como você faria para calcular os lados desse terreno que ela está à procura?

$\boxed{800}$ x

$x + 20$

$x(x+20) = 800$

$x^2 + 20x = 800$

$x^2 + 20x - 800 = 0$

$\Delta = b^2 - 4ac$ $\qquad x = \dfrac{-b \pm \sqrt{\Delta}}{2a}$

$\Delta = 20^2 - 4 \cdot 1 \cdot (-800)$

$\Delta = 400 + 3200$ $\qquad x = \dfrac{-20 \pm 60}{2} \begin{cases} 20 \\ -40 \end{cases}$

$\Delta = 3600$

$\sqrt{\Delta} = \pm 60$

$x = 20 \rightarrow x + 20 = 40 \Rightarrow \boxed{20 \text{ e } 40}$

$x = -40 \rightarrow x + 20 = 20 \Rightarrow \boxed{-40 \text{ e } -20}$

Q9. Baseado na situação apresentada, qual seria o melhor feedback a ser dado para esses alunos?

(a) Parabéns! Você resolveu corretamente a equação de segundo grau necessária para achar a solução do problema. Porém, se o problema pergunta os lados de um terreno, você acha que todos os valores encontrados por você podem ser aceitos?

(b) Parabéns! Sua resposta está quase correta. Porém, os lados do terreno que você apresentou não estão corretos. Você precisa analisar se os valores que você encontrou podem ser considerados como solução do problema.

(c) Parabéns! Eu vejo que você entendeu que era preciso uma equação de segundo grau para resolver o problema. Agora, você precisa reler o problema e analisar as repostas que você deu.

(d) Parabéns! Sua resposta está quase correta. Você só precisa checar a última parte. Se o problema pede os lados de um terreno, você pode aceitar os valores negativos?

A professora Leila então escreveu:

> Parabéns! Você resolveu corretamente a equação de segundo grau necessária para resolver o problema. Porém, se o problema pergunta os lados de um terreno, você acha que todos os valores encontrados por você podem ser aceitos?

Q10. Diante do comentário escrito pela professora Leila, qual seria a melhor atitude que ela deveria tomar após dar esse feedback aos alunos?

(a) Pedir para um aluno que resolveu o problema corretamente, apresentar a solução no quadro para a turma toda.

(b) Pedir para os alunos que não responderam corretamente tentarem resolver novamente, prestando atenção nos comentários que ela fez.

(c) Corrigir o problema no quadro para a classe toda, pedindo que a turma vá dizendo cada passo.

(d) Passar um problema similar para os alunos que não responderam corretamente, pedindo que levem em consideração os comentários que ela fez.

A professora Leila considerou que a melhor ação pedagógica era passar um problema similar para os alunos que não responderam corretamente, pedindo que levassem em consideração os comentários que ela fez.
No final da aula, ela recolheu os problemas e analisou-os novamente. Porém, ela observou que alguns alunos ainda estavam cometendo o mesmo erro.
Conversando com outros professores, eles deram algumas explicações do porquê de isso ter acontecido.

Q11. Na sua opinião, qual professor deu a melhor explicação?

(a) Marcos: "É normal isso acontecer. Alguns alunos têm mais dificuldades que os outros e vão demorar um pouco mais mesmo. Com um pouco mais de treino eles chegam lá."

(b) Lúcia: "Eu acredito que isso tenha acontecido pois o feedback que você escreveu foi muito longo e elaborado. Talvez você devesse ter dado nota para incentivar."

(c) Cris: "Os alunos não conseguiram identificar o erro através do seu comentário. Você deveria ter deixado mais claro que o problema foi considerar as respostas negativas."

(d) Valdir: "Eu acho que você pulou uma etapa. Você deveria primeiro ter dado o mesmo problema para eles refazerem, e depois dar um semelhante para verificar se eles realmente aprenderam."

O professor Luiz chegou um pouco depois e também quis participar da discussão. Ele disse:

> "Em minha opinião, isso aconteceu porque os alunos não estão acostumados a utilizarem o feedback do professor. Um dos papéis do professor é ajudar os alunos a utilizarem essas informações para a melhoria da própria aprendizagem."

Os outros professores acharam o comentário dele bem interessante e quiseram saber um pouco mais como seria uma boa forma de fazer isso.

Q12. Qual seria a melhor forma de responder esses professores?

(a) "Não existe uma boa forma de ensinar isso aos alunos. É melhor dar feedback escrito aos seus alunos com frequência. Quando eles se acostumarem a receber os feedbacks, eles vão acabar desenvolvendo essa habilidade."

(b) "Uma boa forma de fazer isso seria incentivar que os alunos comentassem o trabalho um do outro. Partilhar o feedback com o colega ajudará no desenvolvimento da habilidade de interpretar o feedback do professor."

(c) "Não existe uma boa forma de ensinar isso aos alunos. Ou eles sabem interpretar o feedback ou eles não sabem. Por isso que alguns fizeram o exercício certo na segunda tentativa e outros não."

(d) "Uma boa forma de fazer isso, seria o professor mostrar como ele faria. Ele vai lendo o feedback com os alunos e mostrando a interpretação passo a passo."

Scenario 4

Antes de dar prosseguimento ao conteúdo do 9ºano, a professora Vera decidiu adotar uma estratégia de avaliação entre pares (um aluno avaliando o outro) para avaliar se seus alunos aprenderam a identificar e aplicar as definições e propriedades de raio e/ou diâmetro para calcular comprimento e área.
Para isso, ela elaborou a seguinte atividade:

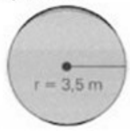
Atividade sobre circunferência e círculo

Nome: ____.

Turma: 9º Ano _____ **Data:** ___/___/_____.

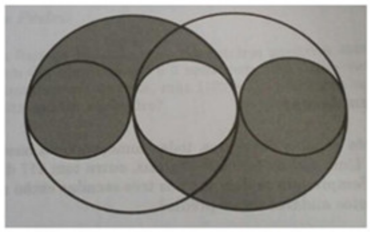
1. Um disco tem diâmetro igual a 11,8 cm. O comprimento de sua circunferência é de aproximadamente:
   a. 3,6 cm
   b. 37,05 cm
   c. 74,1 cm
   d. 118 cm

2. Uma piscina tem formato circular. Sabendo que o seu raio é igual a 3,5 m, pode-se dizer que a área do fundo da piscina é:
   a. 3,8465 m^2
   b. 38,465 m^2
   c. 384,65 m^2
   d. 3846,5 m^2



Utilize $\pi = 3{,}14$

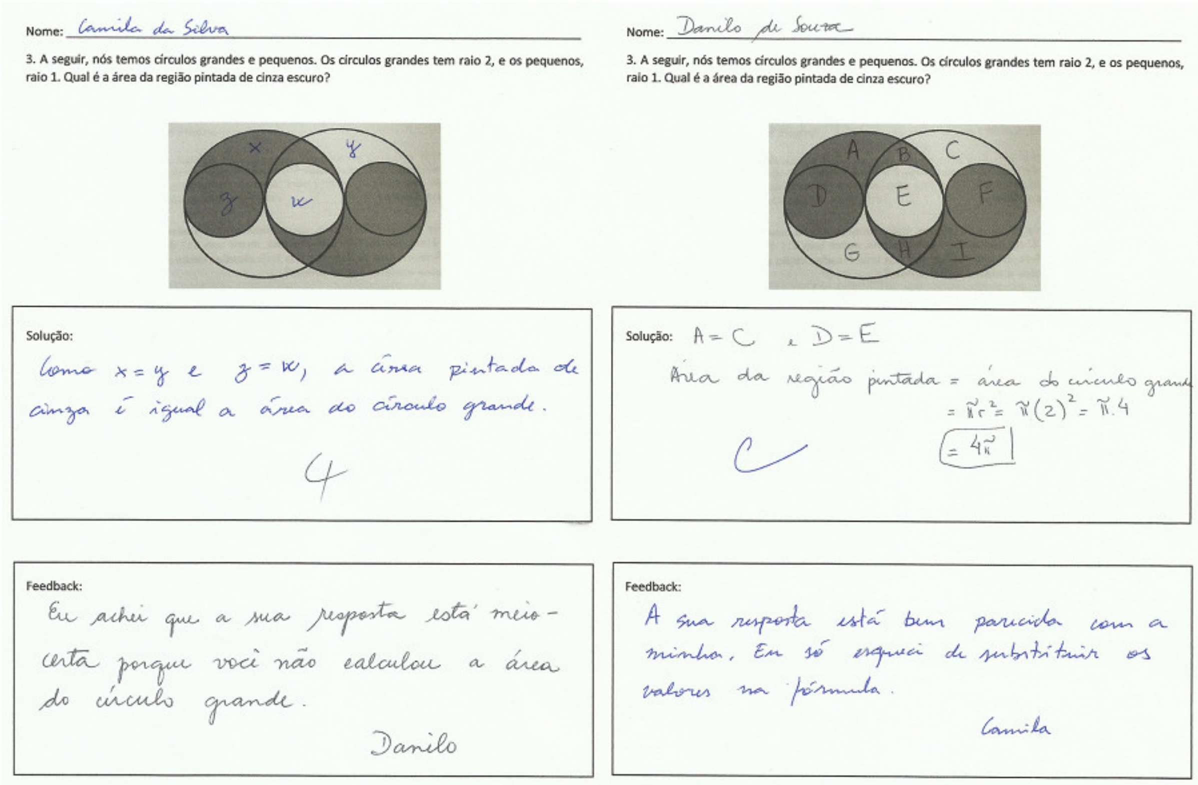
3. Abaixo, temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual a área da região pintada de cinza escuro?



No início da aula, ela entregou a atividade para os alunos e pediu que resolvessem individualmente. Após dar 30 minutos para a resolução, ela deu o seguinte comando:

> "Agora vocês irão trocar a atividade com o colega do lado. Vocês deverão analisar as respostas dadas, decidir se estão certas, meio-certas ou erradas, e ainda escrever um comentário explicando a sua decisão. Em seguida, eu irei recolhê-las para analisá-las durante a coordenação."

Durante sua análise, ela observou comentários bem interessantes, como o da dupla Camila-Danilo:



Nome: Camila da Silva

3. A seguir, nós temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual é a área da região pintada de cinza escuro?

Solução: Como $x = y$ e $z = w$, a área pintada de cinza é igual a área do círculo grande.

4

Feedback: Eu achei que a sua resposta está meio-certa porque você não calculou a área do círculo grande.

Danilo

Nome: Danilo de Souza

3. A seguir, nós temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual é a área da região pintada de cinza escuro?

Solução: $A = C$ e $D = E$

Área da região pintada = área do círculo grande $= \pi r^2 = \pi(2)^2 = \pi . 4$

$\boxed{= 4\pi}$

C

Feedback: A sua resposta está bem parecida com a minha. Eu só esqueci de substituir os valores na fórmula.

Camila

Q13. Analisando a resolução da Camila e o comentário que ela fez na atividade do Danilo, qual a melhor forma de a professora Vera interpretar a aprendizagem da Camila?

(a) Mesmo que sua resposta inicial estivesse incompleta, a Camila aprendeu ao ter a oportunidade de avaliar o Danilo.

(b) A Camila não sabia calcular a área e somente comentou sobre a fórmula porque viu a resolução do Danilo.

(c) Apesar de não ter apresentado os cálculos numéricos, o raciocínio da Camila estava correto.

(d) A Camila sabia calcular a área, porém por algum motivo não aplicou a fórmula.

Na aula seguinte, a professora trouxe todas as atividades de volta. Antes de devolver para os alunos, ela comentou:

"Ontem, eu tive a oportunidade de analisar todas as atividades da aula passada e encontrei resoluções e comentários bem interessantes..."

Q14. Qual seria a melhor forma de orientá-los para que eles utilizem as informações fornecidas pelo colega para a melhoria da sua aprendizagem?

(a) "Agora, eu irei devolvê-las para que vocês analisem o comentário do colega e possam se sair melhor na próxima avaliação".

(b) "Agora, eu irei devolvê-las para que vocês possam refazê-las utilizando as anotações do colega. Se não entenderem ou não concordarem, discutam e tirem dúvidas entre si".

(c) "Agora, eu irei devolvê-las para que vocês possam trocar ideias e comparar suas estratégias de resolução".

(d) "Agora, eu irei devolvê-las para que vocês possam refazê-las analisando o que o colega escreveu. Caso tenham alguma dúvida, me perguntem".

Ao concluir essa etapa, a professora Vera observou que os alunos focaram suas discussões em torno da questão 3. Ela então resolveu levar a discussão para seus colegas ajudarem a avaliar a atividade como um todo.
Veja as contribuições que eles deram:

Ricardo: "Eu acho que a questão 3 gerou mais discussões, pois ela é mais aberta e permite diferentes estratégias de resolução."

Luiza: "Eu já acho que isso aconteceu por causa das questões 1 e 2, pelo fato de elas serem de múltipla-escolha. O aluno vai selecionar a alternativa e pronto. Por isso eu nunca aplico questões de múltipla-escolha com meus alunos."

Ricardo: "Mas as questões 1 e 2 estão condizentes com o objetivo da Vera, de avaliar se o aluno sabe utilizar raio e diâmetro para calcular comprimento e área."

Luiza: "Eu não concordo, pois a 2, na verdade, está avaliando se o aluno sabe o posicionamento da vírgula quando multiplica decimais."

Q15. Qual é a melhor forma de analisarmos essas contribuições?

(a) Ricardo apresentou uma análise relevante, pois abordou tanto a estrutura da questão 3 como os objetivos da professora Vera.

(b) Luiza apresentou uma análise importante, pois ressaltou o problema no formato das questões 1 e 2, o que justifica uma maior atenção à questão 3.

(c) Ambos apresentaram uma análise significativa, pois ressaltaram elementos importantes em suas análises.

(d) Ambos apresentaram uma análise incompleta que foi insuficiente para explicar o problema apresentado.

No final do processo, a professora sentiu necessidade de verificar se os alunos entenderam quais eram os objetivos de aprendizagem tendo como referência as atividades trabalhadas e a forma como os alunos avaliaram uns aos outros.

Q16. Qual seria a melhor intervenção a ser feita pela professora?

(a) "Vocês conseguem me dizer o que estávamos aprendendo sobre circunferência e círculo? Eu vou devolver a atividade para que vocês possam dar uma nova olhada nos exercícios e nos feedbacks e identificar o que aprendemos."

(b) "Vocês se lembram dos objetivos de aprendizagem dessas últimas aulas? Vou escrevê-los aqui no quadro e vou devolver a atividade da aula passada para que vocês identifiquem esses objetivos nos exercícios e nos feedbacks."

(c) "Vocês conseguem me dizer o que é que estávamos aprendendo sobre circunferência e círculo? Eu vou pedir que vocês utilizem a atividade da aula passada e os feedbacks para relembrar. Depois, nós vamos anotar as ideias no quadro."

(d) "Vocês conseguem me dizer qual é o conteúdo que estamos trabalhando nas últimas aulas? Vou entregar a atividade da aula passada para que vocês revisem os exercícios e os feedbacks e tentem identificar quais eram os objetivos de aprendizagem."

Scenario 5

A Gabriela é uma aluna que normalmente se sai muito bem em Matemática. Porém, ela tem uma tendência a começar devagar porque ela demora um pouco para assimilar assuntos novos. A Gabriela tirou 5 no primeiro teste do bimestre e 6 no segundo. O pai dela resolveu conversar com ela pois essas notas estavam abaixo das que ela estava acostumada a tirar.

Pai: "Gabi, eu notei que as suas notas em matemática não estão tão boas como elas costumam ser. O que eu posso fazer para te ajudar, meu amor?"

Gabriela: "Não se preocupe, pai. O professor Cláudio explicou que logo nós teremos uma prova. Essa prova vai ser dividida em seções, sendo que cada seção vai representar esses pequenos testes. Se eu me sair melhor na seção da prova, a minha nota na seção vai substituir a nota do teste."

O pai dela ficou bem feliz em ver a confiança da Gabriela, porém ele não entendeu muito bem porque o professor Cláudio iria substituir as notas. Diante disso, ele resolveu ir à escola conversar com o professor.

Q17. Qual a melhor forma que o professor Cláudio poderia utilizar para explicar ao pai da Gabriela sobre o seu método avaliativo?

(a) "O que eu estou fazendo é o que chamamos recuperação contínua, isto é, nós precisamos dar aos alunos diferentes oportunidades de eles recuperarem as notas ao longo do bimestre."

(b) "Apesar de eu dar uma nota em todos os testes, eles servem mesmo para os alunos praticarem. Se eu não substituir a nota, os alunos não terão motivos para estudar para a prova."

(c) "A nota da prova é o que me interessa mais porque os testes são só para os alunos praticarem. Fazendo isso, a nota da prova é a que vai representar o que eles realmente aprenderam."

(d) "Os testes vão ajudar os alunos a guiarem os estudos e irão ajudá-los a aprender o conteúdo até a prova. A nota então irá melhor refletir a aprendizagem deles."

Q18. E como seria a melhor forma de aconselhar os alunos a utilizarem os testes para a melhoria da aprendizagem?

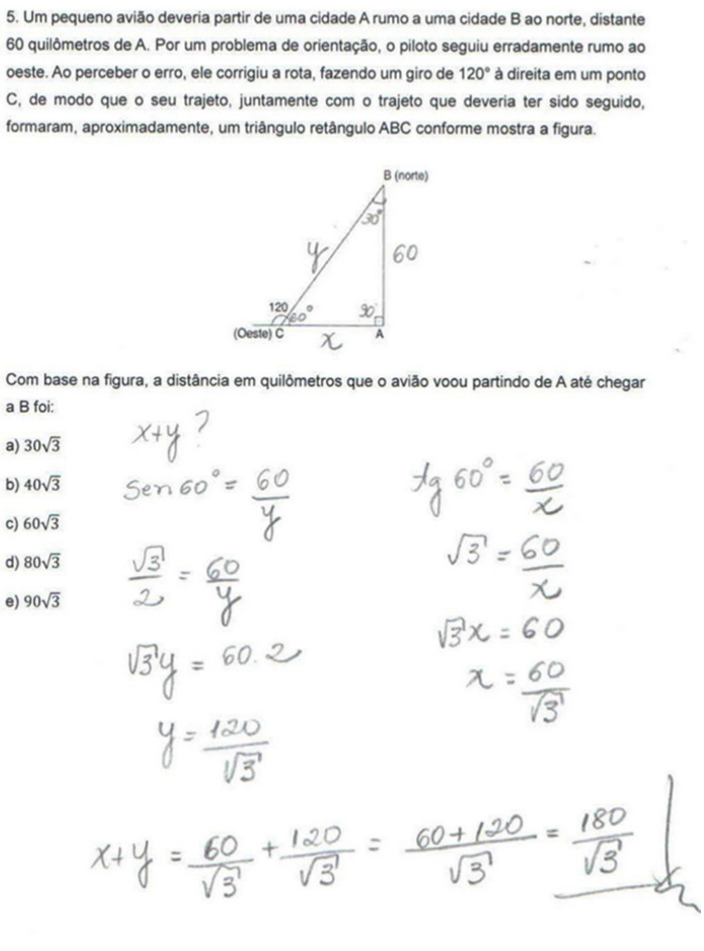
(a) “Vocês devem revisar os testes de forma a identificar o que tiveram dificuldade e trabalhar mais esse conteúdo.”

(b) “Vocês devem procurar ajuda para revisar os testes e aprender como resolver as questões que não conseguiram acertar e ainda refazer aquelas que já acertaram na primeira tentativa.”

(c) “Vocês devem refazer todos os testes várias vezes para praticar para a prova bimestral. Dessa forma, com certeza vocês vão conseguir uma nota melhor.”

(d) “Vocês devem estudar os testes e fazer todos os exercícios para se certificar de que vocês saibam resolver todos.”

Após a prova bimestral, a professora observou que a Gabriela superou as dificuldades apresentadas no primeiro teste, mas ainda poderia melhorar em alguns aspectos do conteúdo do segundo. Veja como ela resolveu uma das questões sobre as relações trigonométricas:


5. Um pequeno avião deveria partir de uma cidade A rumo a uma cidade B ao norte, distante 60 quilômetros de A. Por um problema de orientação, o piloto seguiu erradamente rumo ao oeste. Ao perceber o erro, ele corrigiu a rota, fazendo um giro de 120° à direita em um ponto C, de modo que o seu trajeto, juntamente com o trajeto que deveria ter sido seguido, formaram, aproximadamente, um triângulo retângulo ABC conforme mostra a figura.

Com base na figura, a distância em quilômetros que o avião voou partindo de A até chegar a B foi:

a) $30\sqrt{3}$

b) $40\sqrt{3}$

c) $60\sqrt{3}$

d) $80\sqrt{3}$

e) $90\sqrt{3}$

$x+y$?

$\operatorname{sen} 60° = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60.2$

$y = \frac{120}{\sqrt{3}}$

$tg\, 60° = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$

Diante dessa situação, o professor Cláudio achou que seria uma boa estratégia dar feedback por escrito para ajudá-la a identificar o que ainda ficou faltando.

Q19. Qual seria o melhor feedback a ser dado para a Gabriela?

(a) Muito bem, Gabriela! Seu raciocínio está correto. Agora, você precisa racionalizar o denominador para chegar a uma das alternativas apresentadas.

(b) Muito bem, Gabriela! Você utilizou as relações trigonométricas apropriadamente. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?

(c) Muito bem, Gabriela! Você entendeu que era preciso utilizar as relações trigonométricas para resolver o problema. Agora, você precisa racionalizar o denominador para chegar a uma das alternativas apresentadas.

(d) Muito bem, Gabriela! Sua resolução está quase correta. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?

O professor Cláudio escreveu o seguinte comentário:

> Muito bem, Gabriela! Você utilizou as relações trigonométricas apropriadamente. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?

Ao devolver as provas, ele foi chamando um por um na sua mesa para conversar com eles individualmente.

Q20. O que o professor Cláudio deveria dizer ao entregar a prova comentada para a Gabriela?

(a) “Você está de parabéns, Gabriela, pois você se saiu bem melhor na prova. Você vai ver que eu fiz alguns comentários para você aprender ainda mais.”

(b) “Gabriela, você vai ver que eu escrevi uns comentários em algumas questões, pois você ainda apresentou dificuldade em alguns aspectos. Eu sugiro que você utilize esses comentários para tentar resolver o que você não conseguiu.”

(c) “Muito bem, Gabriela! Você tirou uma boa nota na prova. Agora, eu sugiro que você leia os comentários que eu fiz para tentar resolver as questões que você não conseguiu.”

(d) “Gabriela, você progrediu bastante em relação aos testes, mas em alguns aspectos você ainda pode melhorar. Eu sugiro que você leia o meu comentário para entender o que ficou faltando em algumas questões e tente completá-las.”

Scenario 6

A professora Sônia queria revisar frações e porcentagens com uma de suas turmas do 9º ano. Para isso, ela decidiu pedir para os alunos elaborarem alguns problemas para resolverem juntos em sala de aula. Porém, para que os alunos realizassem a atividade era necessário estabelecer alguns critérios para a elaboração dessas questões.

Q21. Qual seria a melhor forma de definir esses critérios?

(a) Elaborar os critérios com a participação dos alunos.

(b) Definir os critérios e explicá-los detalhadamente aos alunos.

(c) Apresentar alguns critérios e chegar a uma versão final com os alunos.

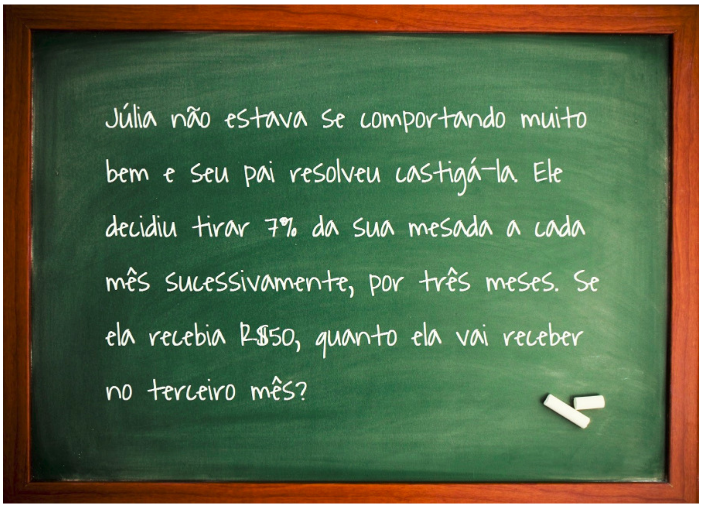
(d) Trazer os critérios e definir valores para cada um deles juntamente com os alunos.

Juntamente com os alunos, eles definiram os seguintes critérios:

Os alunos devem, em duplas, elaborar 4 questões que atendam os seguintes critérios:

1. As questões devem abordar o conteúdo de frações e/ou porcentagens;
2. As questões devem ser contextualizadas;
3. As questões podem ser abertas ou de múltipla-escolha;
4. A dupla deve apresentar pelo menos uma solução para cada questão;

Para a aula seguinte, ela decidiu selecionar algumas dessas questões para passar no quadro para os alunos irem resolvendo oralmente junto com ela. Ela começou com o seguinte problema:



Uma aluna começou a participar:

Ana: "Nós precisamos encontrar quanto é 7% de 50 e daí fazer 3 vezes."

Profa. Sônia: "E o que nós precisamos fazer primeiro?"

Ana: "Tirar de 50."

A professora Sônia anota no quadro dizendo:

> "Isso vai nos dar a quantia no primeiro mês. Nós precisamos fazer o mesmo processo de novo e de novo para encontrarmos a quantia no terceiro mês."

Depois disso, o Paulo entra na discussão.

Paulo: "Ao invés de multiplicar por 0,07 várias vezes, nós poderíamos multiplicar por 0,21 direto."

Q22. Qual é a melhor forma de interpretarmos a resposta do Paulo?

(a) Paulo está seguindo a sugestão da professora e fazendo o mesmo processo de novo e de novo.

(b) Paulo está focando no 0,07 e no 3, sem compreender a situação proposta.

(c) Paulo está multiplicando por 3 ao invés de elevar a 3.

(d) Paulo esqueceu de tirar de 50.

Q23. E nesse caso, qual seria um bom feedback a ser dado para o Paulo?

(a) "Se você multiplicar por 0,21, Paulo, você vai obter 21% de 50, o que não é o que o problema propõe. Você consegue pensar em alguma outra solução?"

(b) "0,21 é 3 vezes 0,07, certo? Você consegue me explicar porque essa solução pode ou não funcionar nesse caso, Paulo?"

(c) "Você consegue me mostrar de onde vem o 0,21, Paulo? Me explica porque você acha que essa solução pode funcionar nesse caso?"

(d) "0,21 pode não funcionar, Paulo. Mas eu vou anotar a sua sugestão no quadro para que a gente possa discutir depois."

E o diálogo continuou da seguinte forma:

Profa. Sônia: "Você consegue me mostrar de onde vem o 0,21, Paulo? Me explica porque você acha que essa solução pode funcionar nesse caso?"

Paulo: "Não são 3 meses? 3 vezes 0,07 é 0,21.

Profa. Sônia: " "Sim, concordo com você: 0,21 é 3 vezes 0,07, mas será que é isso que o problema está perguntando?"

Paulo: "Ué... o problema diz que são 3 meses, por isso que eu achei que tinha que multiplicar por 3."

Profa. Sônia: "São 3 meses, Paulo. Mas são 7% em cada um dos meses, e não tudo no final do terceiro mês, entendeu?"

Paulo: "Ah... entendi. Então tem que tirar 7% primeiro e depois 7% de novo e depois 7% de novo."

Q24. E como a professora poderia dar continuidade de forma a verificar se o Paulo aprendeu com o feedback dado?

(a) "Então vamos fazer o seguinte: resolva no seu caderno e quando eu fizer no quadro, você verifica se obteve o resultado correto."

(b) "Todo mundo entendeu o que o Paulo acabou de dizer? Vamos fazer no quadro juntos?"

(c) "Isso. Você tira 7% de 50. Aí, do que sobrou, você tira mais 7%. E assim por diante. Quer fazer lá no quadro para todo mundo ver?"

(d) "Faz o seguinte então: faz 21% e depois faz desse jeito, tirando 7, 7 e 7. Aí quando você terminar a gente conversa sobre os resultados."

C.2 Final version

This is the final version, in Portuguese, which was sent to FD teachers. The results presented in this study are based on the equivalent of this version in Portuguese.

Scenario 1

O professor João estava trabalhando inequações com seus alunos do 8ºano. Porém, quando ele passou uma atividade avaliativa, ele observou que a maioria dos alunos estava cometendo erros. Veja dois exemplos de como eles estavam resolvendo:

1. $17 - 2x > 25$

$$-2x > 25 - 17$$
$$-2x > 8$$
$$x > -4$$

2. $23 + x < 3x - 9$

$$x - 3x < -9 - 23$$
$$-2x < -32$$
$$x < 16$$

Q1. Qual é a melhor interpretação que pode ser feita em relação à aprendizagem dos alunos, a partir das respostas dadas?

(a) Esses alunos entenderam os passos para resolverem inequações, mas tiveram um erro no finalzinho.

(b) Esses alunos resolveram como se fossem equações, pois não levaram a desigualdade em consideração.

(c) Esses alunos não sabem resolver inequações, pois não obtiveram o resultado correto em nenhum dos dois exercícios.

(d) Esses alunos não prestaram atenção no que o professor ensinou sobre inequações, pois não levaram a desigualdade em consideração.

Q2. Uma vez que a maioria dos alunos apresentou dificuldade, qual seria a melhor forma de dar feedback para os alunos?

(a) Retomar o conteúdo para a turma toda na próxima aula para que eles possam se sair melhor na próxima avaliação.

(b) Entregar a atividade para os alunos e dar a eles outra oportunidade para resolverem os exercícios, pedindo a ajuda do professor quando eles tiverem necessidade.

(c) Reexplicar o conteúdo específico para a turma toda na próxima aula, e dar aos alunos uma outra chance de refazer os exercícios.

(d) Escrever um comentário para os alunos que não resolveram corretamente, e na próxima aula deixar que tentem resolver novamente.

O professor João decidiu que a melhor estratégia era retomar o conteúdo para a turma toda. No início da aula seguinte, ele disse:

> "Nós vamos começar a nossa aula hoje relembrando como se resolve inequações porque eu notei que alguns alunos tiveram dificuldade de resolvê-las na atividade da aula passada."

Q3. Depois que o objetivo estiver claro, qual seria a melhor forma de abordar o conteúdo com os alunos?

(a) Explicar aos os alunos os erros que eles cometeram e devolver a atividade para que eles possam tentar de novo.

(b) Retomar o conteúdo da aula passada e devolver a atividade aos alunos para que eles possam tentar de novo.

(c) Devolver a atividade aos alunos e apresentar a solução correta no quadro comentando os erros que eles cometeram.

(d) Pedir a um aluno que resolveu corretamente para vir ao quadro e demonstrar ao grupo, e em seguida os outros alunos resolvem no próprio caderno.

Scenario 2

O professor André está trabalhando o assunto de divisão por frações com seus alunos do 7^{o} ano. Ele decidiu que na próxima aula avaliará se eles sabem dividir por $\frac{1}{2}$. Para isso, ele elaborou alguns probleminhas:

I. Clara quer dividir uma torta e meia entre dois amigos. Quanto cada amigo vai receber?

II. Você tem R$1,25 e vai logo dobrar essa quantia. Com quanto você vai ficar no final?

III. Sua mãe está fazendo um bolo e a receita pede duas xícaras e $\frac{1}{2}$ de manteiga. Quantos tabletes de manteiga ela vai precisar sendo que cada tablete é equivalente a meia xícara?

Antes de entregar aos seus alunos, ele resolveu te mostrar e perguntou se você acha os probleminhas apropriados para a intenção que ele tem.

Q4. O que você diria ao professor André?

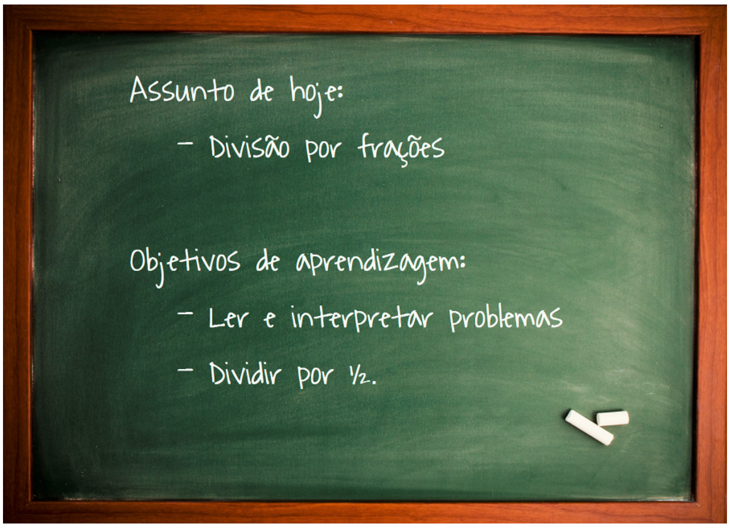
(a) "Eu acredito que você deva revisar sua atividade, pois alguns problemas não estão avaliando o que você quer avaliar."

(b) "Eu acredito que você deva incluir exercícios que exijam que os alunos dividam por outras frações além de $\frac{1}{2}$."

(c) "Eu acredito que a sua atividade esteja apropriada para o seu objetivo e, portanto, está pronta para ser aplicada."

(d) "Eu acredito que você deva incluir mais exercícios, pois três não são suficientes para obter a informação que você está querendo."

Ao começar a aula seguinte, o professor André achou importante deixar claro para os alunos quais eram os objetivos de aprendizagem daquela aula e escreveu no quadro:



Em seguida, ele entregou a atividade aos alunos e colocou-os em trios para que pudessem resolver juntos. Enquanto isso, o professor André ficou andando pela sala tirando dúvidas e observando os alunos durante a resolução da atividade. Ao passar pelo trio Daniel, Caio e Gustavo, ele ouviu a seguinte conversa:

Daniel: "Clara quer dividir uma torta e meia entre dois amigos. Quanto cada amigo vai receber?"

Gustavo: "É só a gente fazer $1 + \frac{1}{2}$ e dividir por $\frac{1}{2}$."

Daniel: "Isso."

Caio: "Não, gente. Ela quer dividir entre 2 amigos. Então tem que fazer $1 + \frac{1}{2}$ e dividir por 2."

Daniel: "Mas está escrito no quadro que os exercícios são para dividir por meio. Então a gente tem que dividir por meio."

Caio: "Mas também tem o outro objetivo que é ler e interpretar problemas, então eu acho que é uma pegadinha."

Gustavo: "Tá bom... Mas mesmo assim. Se ela quer dividir entre dois amigos, ela quer dividir na metade, isto é, dividir por meio, que é o que eu tinha dito desde o início."

Q5. Baseado na conversa entre os alunos, o que o professor André pode concluir sobre a aprendizagem do Gustavo?

(a) O Gustavo está dividindo por meio, pois está com dificuldade de interpretar o problema.

(b) O Gustavo está considerando dividir por meio e dividir no meio como sendo a mesma coisa.

(c) O Gustavo está dividindo por meio por causa dos objetivos de aprendizagem escritos no quadro.

(d) Como o problema não aborda divisão por meio, pouco se pode concluir sobre a aprendizagem do Gustavo.

O professor então percebeu que estava havendo uma confusão. Porém, ele ficou em dúvida se era porque os alunos estavam confundindo dividir por meio com dividir no meio, ou se era porque estavam sendo influenciados pelos objetivos escritos no quadro no início da aula.

Q6. Como o professor poderia ter evitado essa confusão ao lidar com os objetivos?

(a) Ele deveria ter utilizado alguma estratégia para verificar se os alunos haviam entendido os objetivos.

(b) Ele deveria ter escrito apenas um objetivo no quadro para que ficasse mais claro para os alunos.

(c) Ele deveria ter falado os objetivos ao invés de escrevê-los para ficar mais claro para os alunos.

(d) Ele deveria ter escrito mais detalhes para que os objetivos ficassem mais claros para os alunos.

Diante dessa situação, o professor André decidiu intervir na conversa.

Prof. André: "E você, Daniel? O que você acha?"

Daniel: "Eu acho que a gente tem que dividir por meio porque dividir entre dois amigos é dividir no meio..."

Prof. André: "Então você concorda com o Gustavo... E você, Caio? Concorda com seus colegas?"

Caio: "Não. Eu ainda acho que se a Clara quer dividir entre dois amigos, nós temos que dividir por dois."

Q7. Como o professor André poderia dar prosseguimento de forma a ajudar o grupo a avaliar qual das formas é a mais apropriada?

(a) "Acho que vocês chegaram num impasse. Vocês precisam conversar um pouco mais para chegarem a uma decisão."

(b) "Vamos ler o problema novamente? Podemos dizer que dividir no meio e dividir por meio é a mesma coisa?"

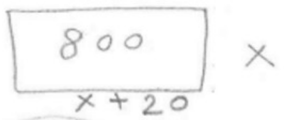
(c) "E porque vocês não fazem as duas contas e analisam as respostas de acordo com o que o problema sugere?"

(d) "Vamos trazer a discussão para a turma toda e ver se conseguimos chegar a uma conclusão?"

Scenario 3

Durante a última semana, a professora Leila estava trabalhando equações de 2º grau com seus alunos do 9º ano. Na última aula, ela decidiu passar alguns problemas para avaliar se seus alunos haviam aprendido a verificar se as raízes da equação poderiam ser aceitas como solução do problema. Ela observou que alguns poucos alunos estavam respondendo como o exemplo abaixo.

Dona Eliana acabou de descobrir que está grávida do seu terceiro filho e decide construir uma casa maior. Ela ficou bem interessada em alguns terrenos de 800m² que encontrou nos classificados do jornal. Porém, baseada no projeto que ela tem para a nova casa, seria necessário um terreno retangular em que um dos lados fosse 20m maior do que o outro. Ela então resolve pedir sua ajuda. Como você faria para calcular os lados desse terreno que ela está à procura?



$$x(x+20) = 800$$
$$x^2 + 20x = 800$$
$$x^2 + 20x - 800 = 0$$

$$\Delta = b^2 - 4ac \qquad x = \frac{-b \pm \sqrt{\Delta}}{2a}$$
$$\Delta = 20^2 - 4 \cdot 1 \cdot (-800)$$
$$\Delta = 400 + 3200 \qquad x = \frac{-20 \pm 60}{2} \begin{cases} 20 \\ -40 \end{cases}$$
$$\Delta = 3600$$
$$\sqrt{\Delta} = \pm 60$$

$$x = 20 \rightarrow x + 20 = 40 \Rightarrow \boxed{20 \text{ e } 40}$$
$$x = -40 \rightarrow x + 20 = 20 \Rightarrow \boxed{-40 \text{ e } -20}$$

Q8. Baseado na situação apresentada, qual seria o melhor feedback a ser dado para esses alunos?

(a) Parabéns! Você resolveu corretamente a equação de segundo grau necessária para achar a solução do problema. Porém, se o problema pergunta os lados de um terreno, você acha que todos os valores encontrados por você podem ser aceitos?

(b) Parabéns! Sua resposta está quase correta. Porém, os lados do terreno que você apresentou não estão corretos. Você precisa analisar se os valores que você encontrou podem ser considerados como solução do problema.

(c) Parabéns! Eu vejo que você entendeu que era preciso uma equação de segundo grau para resolver o problema. Agora, você precisa reler o problema e analisar as repostas que você deu.

(d) Parabéns! Sua resposta está quase correta. Você só precisa checar a última parte. Se o problema pede os lados de um terreno, você pode aceitar os valores negativos?

A professora Leila então escreveu:

> "Parabéns! Você resolveu corretamente a equação de segundo grau necessária para achar a solução do problema. Porém, se o problema pergunta os lados de um terreno, você acha que todos os valores encontrados por você podem ser aceitos?"

Q9. Diante do comentário escrito pela professora Leila, qual seria a melhor atitude que ela deveria tomar após dar esse feedback aos alunos?

(a) Pedir para um aluno que resolveu o problema corretamente, apresentar a solução no quadro para a turma toda.

(b) Pedir para os alunos que não responderam corretamente tentarem resolver novamente, prestando atenção nos comentários que ela fez.

(c) Corrigir o problema no quadro para a classe toda, pedindo que a turma vá dizendo cada passo.

(d) Passar um problema similar para os alunos que não responderam corretamente, pedindo que levem em consideração os comentários que ela fez.

Scenario 4

Antes de dar prosseguimento ao conteúdo do 9ºano, a professora Vera decidiu adotar uma estratégia de avaliação entre pares (um aluno avaliando o outro) para avaliar se seus alunos aprenderam a identificar e aplicar as definições e propriedades de raio e/ou diâmetro para calcular comprimento e área.
Para isso, ela elaborou a seguinte atividade:

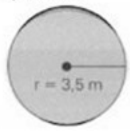
Atividade sobre circunferência e círculo

Nome: ________________________________.

Turma: 9º Ano ______ **Data:** ___/___/_____.

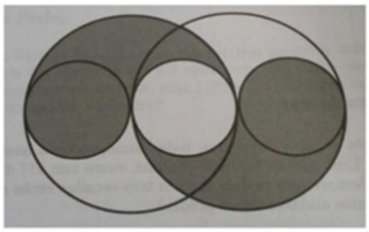
1. Um disco tem diâmetro igual a 11,8 cm. O comprimento de sua circunferência é de aproximadamente:
   a. 3,6 cm
   b. 37,05 cm
   c. 74,1 cm
   d. 118 cm

2. Uma piscina tem formato circular. Sabendo que o seu raio é igual a 3,5 m, pode-se dizer que a área do fundo da piscina é:
   a. 3,8465 m²
   b. 38,465 m²
   c. 384,65 m²
   d. 3846,5 m²



Utilize π = 3,14

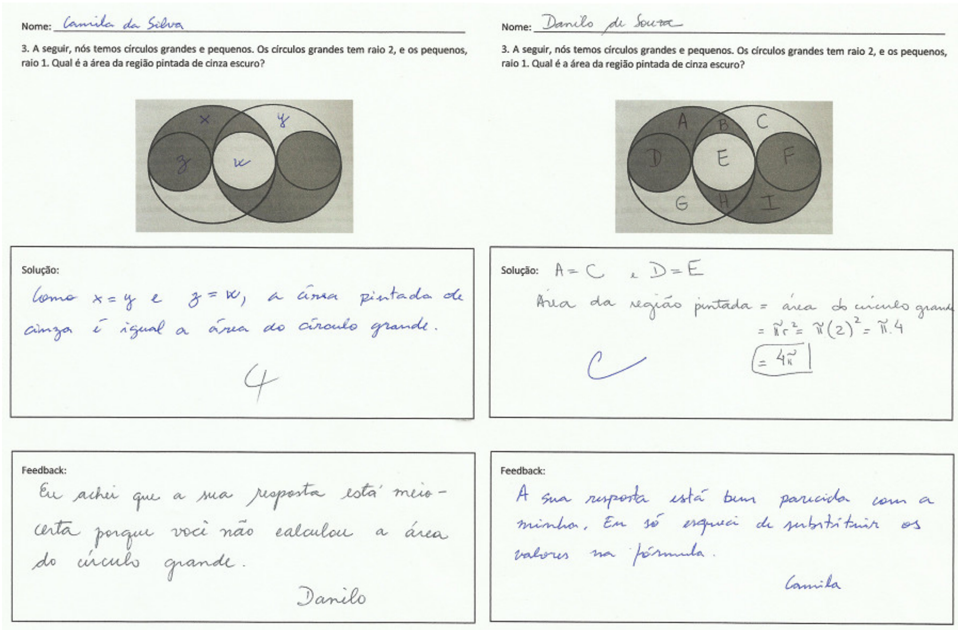
3. Abaixo, temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual a área da região pintada de cinza escuro?



No início da aula, ela entregou a atividade para os alunos e pediu que resolvessem individualmente. Quando eles terminaram ela deu o seguinte comando:

> "Agora vocês irão trocar a atividade com o colega do lado. Vocês deverão analisar as respostas dadas, decidir se estão certas, meio-certas ou erradas, e ainda escrever um comentário explicando a sua decisão. Em seguida, eu irei recolhê-las para analisá-las durante a coordenação."

Durante sua análise, ela observou comentários bem interessantes, como o da dupla Camila-Danilo:


Nome: Camila da Silva

3. A seguir, nós temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual é a área da região pintada de cinza escuro?

x y z w

Solução:
Como x = y e z = w, a área pintada de cinza é igual a área do círculo grande.

4

Feedback:
Eu achei que a sua resposta está meio-certa porque você não calculou a área do círculo grande.

Danilo

Nome: Danilo de Souza

3. A seguir, nós temos círculos grandes e pequenos. Os círculos grandes tem raio 2, e os pequenos, raio 1. Qual é a área da região pintada de cinza escuro?

A B C D E F G H I

Solução: A = C e D = E
Área da região pintada = área do círculo grande
$= \pi r^2 = \pi(2)^2 = \pi . 4$
$\boxed{= 4\pi}$

C

Feedback:
A sua resposta está bem parecida com a minha. Eu só esqueci de substituir os valores na fórmula.

Camila

Q10. Analisando a resolução da Camila e o comentário que ela fez na atividade do Danilo, qual a melhor forma de a professora Vera interpretar a aprendizagem da Camila?

(a) Mesmo que sua resposta inicial estivesse incompleta, a Camila aprendeu ao ter a oportunidade de avaliar o Danilo.

(b) A Camila não sabia calcular a área e somente comentou sobre a fórmula porque viu a resolução do Danilo.

(c) Apesar de não ter apresentado os cálculos numéricos, o raciocínio da Camila estava correto.

(d) A Camila sabia calcular a área, porém por algum motivo não aplicou a fórmula.

Na aula seguinte, a professora trouxe todas as atividades de volta. Antes de devolver para os alunos, ela comentou:

> "Ontem, eu tive a oportunidade de analisar todas as atividades da aula passada e encontrei resoluções e comentários bem interessantes. Agora, eu

irei devolvê-las para que vocês possam refazê-las utilizando as anotações do colega. Se não entenderem ou não concordarem, discutam e tirem dúvidas entre si".

Ela então retornou as atividades e deixou que eles trocassem ideias entre si.
Ao concluir essa etapa, a professora Vera observou que os alunos focaram suas discussões em torno da questão 3. Ela então resolveu levar a discussão para seus colegas ajudarem a avaliar a atividade como um todo.
Veja as contribuições que eles deram:

Ricardo: "Eu acho que a questão 3 gerou mais discussões, pois ela é mais aberta e permite diferentes estratégias de resolução."

Luiza: "Eu já acho que isso aconteceu por causa das questões 1 e 2, pelo fato de elas serem de múltipla-escolha. O aluno vai selecionar a alternativa e pronto. Por isso eu nunca aplico questões de múltipla-escolha com meus alunos."

Ricardo: "Mas as questões 1 e 2 estão condizentes com o objetivo da Vera, de avaliar se o aluno sabe utilizar raio e diâmetro para calcular comprimento e área."

Luiza: "Eu não concordo, pois a 2, na verdade, está avaliando se o aluno sabe o posicionamento da vírgula quando multiplica decimais."

Q11. Qual é a melhor forma de analisarmos essas contribuições?

(a) Ricardo apresentou uma análise relevante, pois abordou tanto a estrutura da questão 3 como os objetivos da professora Vera.

(b) Luiza apresentou uma análise importante, pois ressaltou o problema no formato das questões 1 e 2, o que justifica uma maior atenção à questão 3.

(c) Ambos apresentaram uma análise significativa, pois ressaltaram elementos importantes em suas análises.

(d) Ambos apresentaram uma análise incompleta que foi insuficiente para explicar o problema apresentado.

No final do processo, a professora sentiu necessidade de verificar se os alunos entenderam quais eram os objetivos de aprendizagem tendo como referência as atividades trabalhadas e a forma como os alunos avaliaram uns aos outros.

Q12. Qual seria a melhor intervenção a ser feita pela professora?

(a) “Vocês conseguem me dizer o que estávamos aprendendo sobre circunferência e círculo? Eu vou devolver a atividade para que vocês possam dar uma nova olhada nos exercícios e nos feedbacks e identificar o que aprendemos.”

(b) “Vocês se lembram dos objetivos de aprendizagem dessas últimas aulas? Vou escrevê-los aqui no quadro e vou devolver a atividade da aula passada para que vocês identifiquem esses objetivos nos exercícios e nos feedbacks.”

(c) “Vocês conseguem me dizer o que é que estávamos aprendendo sobre circunferência e círculo? Eu vou pedir que vocês utilizem a atividade da aula passada e os feedbacks para relembrar. Depois, nós vamos anotar as ideias no quadro.”

(d) “Vocês conseguem me dizer qual é o conteúdo que a gente vem trabalhando nas últimas aulas? Vou entregar a atividade da aula passada para que vocês revisem os exercícios e os feedbacks e tentem identificar quais eram os objetivos de aprendizagem.”

Scenario 5

A Gabriela é uma aluna que normalmente se sai muito bem em Matemática. Porém, ela tem uma tendência a começar devagar porque ela demora um pouco para assimilar assuntos novos. A Gabriela tirou 5 no primeiro teste do bimestre e 6 no segundo. O pai dela resolveu conversar com ela pois essas notas estavam abaixo das que ela estava acostumada a tirar.

Pai: "Gabi, eu notei que as suas notas em matemática não estão tão boas como elas costumam ser. O que eu posso fazer para te ajudar, meu amor?"

Gabriela: "Não se preocupe, pai. O professor Cláudio explicou que logo nós teremos uma prova. Essa prova vai ser dividida em seções, sendo que cada seção vai representar esses pequenos testes. Se eu me sair melhor na seção da prova, a minha nota na seção vai substituir a nota do teste."

O pai dela ficou bem feliz em ver a confiança da Gabriela, porém ele não entendeu muito bem porque o professor Cláudio iria substituir as notas. Diante disso, ele resolveu ir à escola conversar com o professor.

Q13. Qual a melhor forma que o professor Cláudio poderia utilizar para explicar ao pai da Gabriela sobre o seu método avaliativo?

(a) "O que eu estou fazendo é o que chamamos recuperação contínua, isto é, nós precisamos dar aos alunos diferentes oportunidades de eles recuperarem as notas ao longo do bimestre."

(b) "Apesar de eu dar uma nota em todos os testes, eles servem mesmo para os alunos praticarem. Se eu não substituir a nota, os alunos não terão motivos para estudar para a prova."

(c) "A nota da prova é o que me interessa mais porque os testes são só para os alunos praticarem. Fazendo isso, a nota da prova é a que vai representar o que eles realmente aprenderam."

(d) "Os testes vão ajudar os alunos a guiarem os estudos e irão ajudá-los a aprender o conteúdo até a prova. A nota então irá melhor refletir a aprendizagem deles."

Q14. E como seria a melhor forma de aconselhar os alunos a utilizarem os testes para a melhoria da aprendizagem?

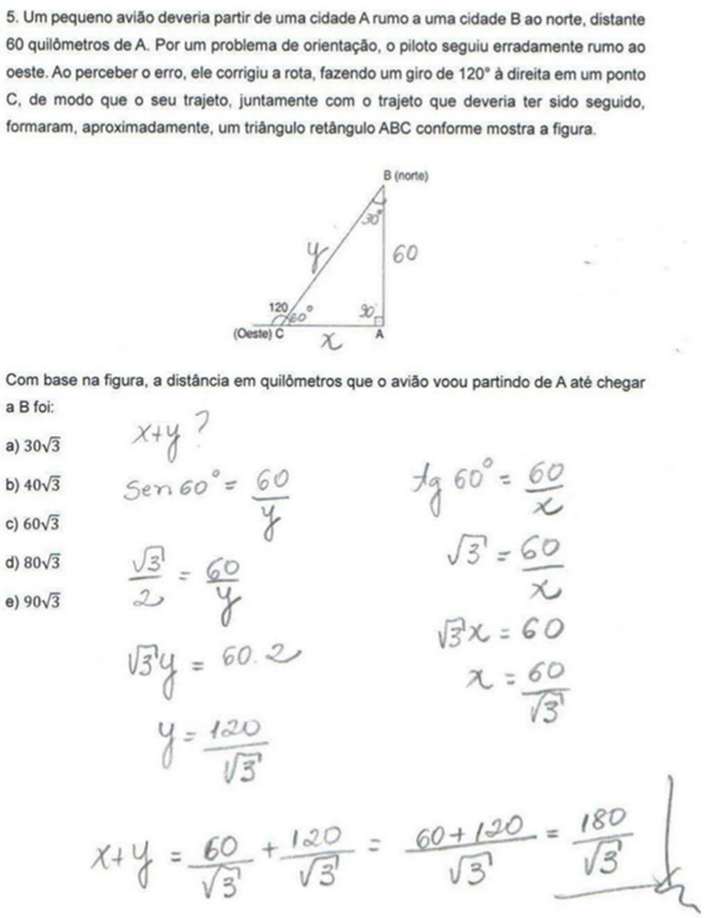
(a) “Vocês devem revisar os testes de forma a identificar o que tiveram dificuldade e trabalhar mais esse conteúdo.”

(b) “Vocês devem procurar ajuda para revisar os testes e aprender como resolver as questões que não conseguiram acertar e ainda refazer aquelas que já acertaram na primeira tentativa.”

(c) “Vocês devem refazer todos os testes várias vezes para praticar para a prova bimestral. Dessa forma, com certeza vocês vão conseguir uma nota melhor.”

(d) “Vocês devem estudar os testes e fazer todos os exercícios para se certificar de que vocês saibam resolver todos.”

Após a prova bimestral, a professora observou que a Gabriela superou as dificuldades apresentadas no primeiro teste, mas ainda poderia melhorar em alguns aspectos do conteúdo do segundo. Veja como ela resolveu uma das questões sobre as relações trigonométricas:


5. Um pequeno avião deveria partir de uma cidade A rumo a uma cidade B ao norte, distante 60 quilômetros de A. Por um problema de orientação, o piloto seguiu erradamente rumo ao oeste. Ao perceber o erro, ele corrigiu a rota, fazendo um giro de 120° à direita em um ponto C, de modo que o seu trajeto, juntamente com o trajeto que deveria ter sido seguido, formaram, aproximadamente, um triângulo retângulo ABC conforme mostra a figura.

Com base na figura, a distância em quilômetros que o avião voou partindo de A até chegar a B foi:

a) $30\sqrt{3}$

b) $40\sqrt{3}$

c) $60\sqrt{3}$

d) $80\sqrt{3}$

e) $90\sqrt{3}$

$x+y$?

$\text{Sen}\,60° = \frac{60}{y}$

$\frac{\sqrt{3}}{2} = \frac{60}{y}$

$\sqrt{3}y = 60.2$

$y = \frac{120}{\sqrt{3}}$

$\text{tg}\,60° = \frac{60}{x}$

$\sqrt{3} = \frac{60}{x}$

$\sqrt{3}x = 60$

$x = \frac{60}{\sqrt{3}}$

$$x+y = \frac{60}{\sqrt{3}} + \frac{120}{\sqrt{3}} = \frac{60+120}{\sqrt{3}} = \frac{180}{\sqrt{3}}$$

Diante dessa situação, o professor Cláudio achou que seria uma boa estratégia dar feedback por escrito para ajudá-la a identificar o que ainda ficou faltando.

Q15. Qual seria o melhor feedback a ser dado para a Gabriela?

(a) Muito bem, Gabriela! Seu raciocínio está correto. Agora, você precisa racionalizar o denominador para chegar a uma das alternativas apresentadas.

(b) Muito bem, Gabriela! Você entendeu o problema e utilizou as relações trigonométricas apropriadamente. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?

(c) Muito bem, Gabriela! Você entendeu que era preciso utilizar as relações trigonométricas para resolver o problema. Agora, você precisa racionalizar o denominador para chegar a uma das alternativas apresentadas.

(d) Muito bem, Gabriela! Sua resolução está quase correta. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?

O professor Cláudio escreveu o seguinte comentário:

> “Muito bem, Gabriela! Você entendeu o problema e utilizou as relações trigonométricas apropriadamente. Você consegue pensar o que mais você precisa fazer para encontrar uma das alternativas apresentadas?”

Ao devolver as provas, ele foi chamando um por um na sua mesa para conversar com eles individualmente.

Q16. O que o professor Cláudio deveria dizer ao entregar a prova comentada para a Gabriela?

(a) “Você está de parabéns, Gabriela, pois você se saiu bem melhor na prova. Você vai ver que eu fiz alguns comentários para você aprender ainda mais.”

(b) “Gabriela, você vai ver que eu escrevi uns comentários em algumas questões, pois você ainda apresentou dificuldade em alguns aspectos. Eu sugiro que você utilize esses comentários para tentar resolver o que você não conseguiu.”

(c) “Muito bem, Gabriela! Você tirou uma boa nota na prova. Agora, eu sugiro que você leia os comentários que eu fiz para tentar resolver as questões que você não conseguiu.”

(d) “Gabriela, você progrediu bastante em relação aos testes, mas em alguns aspectos você ainda pode melhorar. Eu sugiro que você leia o meu comentário para entender o que ficou faltando em algumas questões e tente completá-las.”

Scenario 6

A professora Sônia queria revisar frações e porcentagens com uma de suas turmas do 9º ano. Para isso, ela decidiu pedir para os alunos elaborarem alguns problemas para resolverem juntos em sala de aula. Porém, para que os alunos realizassem a atividade era necessário estabelecer alguns critérios para a elaboração dessas questões.

Q17. Qual seria a melhor forma de definir esses critérios?

(a) Elaborar os critérios com a participação dos alunos.

(b) Definir os critérios e explicá-los detalhadamente aos alunos.

(c) Apresentar alguns critérios e chegar a uma versão final com os alunos.

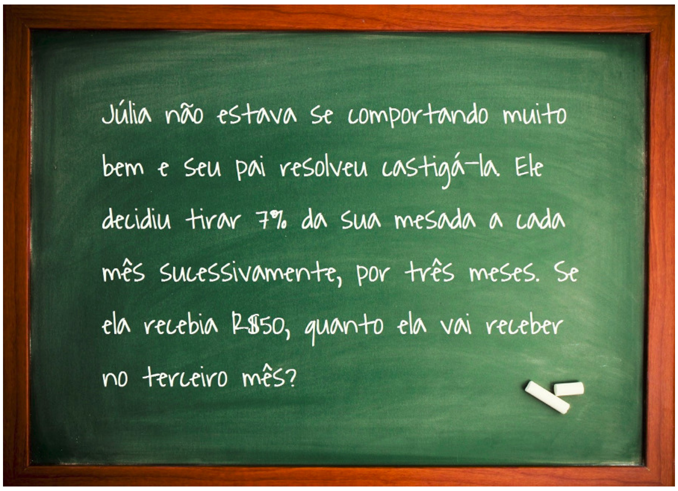
(d) Trazer os critérios e definir valores para cada um deles juntamente com os alunos.

Juntamente com os alunos, eles definiram os seguintes critérios:

Os alunos devem, em duplas, elaborar 4 questões que atendam os seguintes critérios:

1. As questões devem abordar o conteúdo de frações e/ou porcentagens;
2. As questões devem ser contextualizadas;
3. As questões podem ser abertas ou de múltipla-escolha;
4. A dupla deve apresentar pelo menos uma solução para cada questão;

Para a aula seguinte, ela decidiu selecionar algumas dessas questões para passar no quadro para os alunos resolverem oralmente junto com ela. Ela começou com o seguinte problema:



Uma aluna começou a participar:

Ana: "Nós precisamos encontrar quanto é 7% de 50 e daí fazer 3 vezes."

Profa. Sônia: "E o que nós precisamos fazer primeiro?"

Ana: "Tirar de 50."

A professora Sônia anota no quadro dizendo:

> "Isso vai nos dar a quantia no primeiro mês. Nós precisamos fazer o mesmo processo de novo e de novo para encontrarmos a quantia no terceiro mês."

Depois disso, o Paulo entra na discussão.

Paulo: "Ao invés de multiplicar por 0,07 várias vezes, nós poderíamos multiplicar por 0,21 direto."

Q18. Qual é a melhor forma de interpretarmos a resposta do Paulo?

(a) Paulo está seguindo a sugestão da professora e fazendo o mesmo processo de novo e de novo.

(b) Paulo está focando no 0,07 e no 3, sem compreender a situação proposta.

(c) Paulo está multiplicando por 3 ao invés de elevar a 3.

(d) Paulo esqueceu de tirar de 50.

Q19. E nesse caso, qual seria um bom feedback a ser dado para o Paulo?

(a) "Se você multiplicar por 0,21, Paulo, você vai obter 21% de 50, o que não é o que o problema propõe. Você consegue pensar em alguma outra solução?"

(b) "0,21 é 3 vezes 0,07, certo? Você consegue me explicar porque essa solução pode ou não funcionar nesse caso, Paulo?"

(c) "Você consegue me mostrar de onde vem o 0,21, Paulo? Me explica porque você acha que essa solução pode funcionar nesse caso?"

(d) "0,21 pode não funcionar, Paulo. Mas eu vou anotar a sua sugestão no quadro para que a gente possa discutir depois."

E o diálogo continuou da seguinte forma:

Profa. Sônia: "Você consegue me mostrar de onde vem o 0,21, Paulo? Me explica porque você acha que essa solução pode funcionar nesse caso?"

Paulo: "Não são 3 meses? 3 vezes 0,07 é 0,21.

Profa. Sônia: " "Sim, concordo com você: 0,21 é 3 vezes 0,07, mas será que é isso que o problema está perguntando?"

Paulo: "Ué... o problema diz que são 3 meses, por isso que eu achei que tinha que multiplicar por 3."

Profa. Sônia: "São 3 meses, Paulo. Mas são 7% em cada um dos meses, e não tudo no final do terceiro mês, entendeu?"

Paulo: "Ah... entendi. Então tem que tirar 7% primeiro e depois 7% de novo e depois 7% de novo."

Q20. E como a professora poderia dar continuidade de forma a verificar se o Paulo aprendeu com o feedback dado?

(a) "Então vamos fazer o seguinte: resolva no seu caderno e quando eu fizer no quadro, você verifica se obteve o resultado correto."

(b) "Todo mundo entendeu o que o Paulo acabou de dizer? Vamos fazer no quadro juntos?"

(c) "Isso. Você tira 7% de 50. Aí, do que sobrou, você tira mais 7%. E assim por diante. Quer fazer lá no quadro para todo mundo ver?"

(d) "Faz o seguinte então: faz 21% e depois faz desse jeito, tirando 7, 7 e 7. Aí quando você terminar a gente conversa sobre os resultados."

D Additional information

D.1 Results of SUB-PHASE 2A

The ranking presented below are in reference to the version in Appendix B.1.

Table 11.4: Researchers' ranking on first time of SUB-PHASE 2A

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Question 1										
B	A	B	B	B	A	B	B	B	B	B
A	B	C	C	C	C	C	A	A	A	A
C	C	A	A	A	B	A	C	C	C	C
D	D	D	D	D	D	D	D	D	D	D
Question 2										
B	B	D	A	D	A	B	A	A	B	D
C	A	B	B	B	B	A	B	B	A	B
D	D	A	C	A	D	D	C	D	D	A
A	C	C	D	C	C	C	D	C	C	C
Question 3										
D	A	D	A	D	D	A	D	D	A	A
A	C	A	D	A	A	D	A	A	D	D
B	D	C	B	C	C	C	C	B	B	B
C	B	B	C	B	B	B	B	C	C	C
Question 4										
A	C	B	B	B	B	B	A	C	B	A
D	A	A	A	C	A	D	C	B	A	B
C	D	D	D	A	D	C	D	D	D	C
B	B	C	C	D	C	A	B	A	C	D
Question 5										
C	C	C	D	C	D	C	C	C	D	C
D	D	D	C	D	C	D	D	D	C	D

A	A	A	B	A	B	A	A	B	A	A
B	B	B	A	B	A	B	B	A	B	B
Question 6										
D	B	D	D	D	D	B	D	D	B	D
B	A	B	B	B	B	A	B	B	A	B
C	C	A	C	A	C	C	C	A	D	A
A	D	C	A	C	A	D	A	C	C	C
Question 7										
B	A	B	B	B	C	D	B	B	B	B
A	C	C	C	A	B	B	A	A	C	A
D	B	D	A	C	A	A	C	C	A	D
C	D	A	D	D	D	C	D	D	D	C
Question 8										
D	C	D	C	C	C	A	D	A	C	C
C	A	A	B	A	D	D	A	D	A	A
B	B	C	D	D	A	C	B	C	B	B
A	D	B	A	B	B	B	C	B	D	D
Question 9										
B	D	D	D	D	D	D	D	D	D	A
D	A	C	B	C	C	A	C	A	A	D
C	B	A	A	A	A	B	B	C	B	C
A	C	B	C	B	B	C	A	B	C	B
Question 10										
C	A	C	B	C	C	B	C	B	C	C
B	B	B	C	B	B	C	D	C	A	A
A	D	D	A	A	A	D	A	D	B	B
D	C	A	D	D	D	A	B	A	D	D
Question 11										
B	A	A	A	A	A	A	A	C	A	A
D	D	B	D	C	B	B	B	B	B	C

C	B	D	B	D	D	C	C	A	C	D
A	C	C	C	B	C	D	D	D	D	B
Question 12										
A	D	D	D	D	D	A	A	D	D	D
D	A	A	A	A	A	D	D	B	A	A
B	C	B	B	B	B	C	C	A	B	B
C	B	C	C	C	C	B	B	C	C	C
Question 13										
A	A	A	C	A	A	A	A	B	A	B
D	B	B	A	B	B	C	C	A	C	A
B	C	C	B	C	C	D	D	C	B	C
C	D	D	D	D	D	B	B	D	D	D

The ranking presented below are in reference to the version in Appendix B.3.

Table 11.5: Researchers' ranking on second time of SUB-PHASE 2A

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Question 1										
B	B	B	B	B	A	B	B	B	B	B
A	A	C	C	C	C	C	A	A	A	A
C	C	A	A	A	B	A	C	C	C	C
D	D	D	D	D	D	D	D	D	D	D
Question 2										
B	B	A	A	D	A	B	D	A	B	A
D	A	B	B	B	B	A	B	B	A	B
A	D	D	C	A	D	D	A	D	D	C
C	C	C	D	C	C	C	C	C	C	D
Question 3										

D	A	D	D	A	A	D	D	D	A	A
A	C	A	A	D	D	A	A	A	D	D
B	D	C	C	B	C	C	C	B	B	B
C	B	B	B	C	B	B	B	C	C	C
Question 4										
A	C	B	B	B	B	A	A	A	B	A
D	A	A	A	C	A	D	C	C	A	B
C	D	D	D	A	D	C	D	D	D	C
B	B	C	C	D	C	B	B	B	C	D
Question 5										
D	D	D	D	B	D	B	D	B	D	B
B	B	B	B	D	B	D	B	D	B	D
A	A	A	C	A	C	A	A	C	A	A
C	C	C	A	C	A	C	C	A	C	C
Question 6										
A	B	A	A	A	A	B	B	A	B	A
B	A	B	B	B	B	A	A	B	A	B
C	C	C	C	C	C	C	C	D	D	D
D	D	D	D	D	D	D	D	C	C	C
Question 7										
C	A	A	A	A	A	A	A	C	C	C
A	C	C	C	C	C	C	C	A	A	A
D	D	D	D	D	B	D	B	B	B	D
B	B	B	B	B	D	B	D	D	D	B
Question 8										
D	D	D	A	A	A	A	D	A	D	D
A	A	A	D	D	D	D	A	D	A	A
B	B	C	C	C	C	C	B	C	B	B
C	C	B	B	B	B	B	C	B	C	C

Question 9

B	D	D	D	B	B	B	D	D	B	B
D	B	B	B	D	D	D	B	B	D	D
C	A	A	A	A	A	A	C	C	A	C
A	C	C	C	C	C	C	A	A	C	A

Question 10

C	C	C	B	C	C	B	C	B	C	C
B	B	B	C	B	B	C	B	C	B	B
A	D	D	A	A	A	D	A	D	A	A
D	A	A	D	D	D	A	D	A	D	D

Question 11

C	C	D	C	D	D	D	D	C	D	D
D	D	C	D	C	C	C	C	D	C	C
A	A	A	A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B	B	B	B

Question 12

C	D	D	D	D	D	C	C	D	D	D
D	C	C	C	C	C	D	D	C	C	C
B	A	B	B	B	B	A	A	A	B	B
A	B	A	A	A	A	B	B	B	A	A

Question 13

A	A	A	C	A	A	A	A	C	A	C
C	C	C	A	C	C	C	C	A	C	A
B	B	B	B	B	B	D	D	B	B	B
D	D	D	D	D	D	B	B	D	D	D

Question 14

B	B	B	A	A	A	B	B	B	B	B
A	A	A	B	B	B	A	A	A	A	A
D	C	C	C	C	C	D	D	C	C	C

C	D	D	D	D	D	C	C	D	D	D
Question 15										
C	B	C	C	C	C	C	B	C	B	C
B	C	B	B	B	B	B	C	B	C	B
D	D	D	D	D	D	D	A	D	A	D
A	A	A	A	A	A	A	D	A	D	A
Question 16										
C	C	D	C	C	C	D	D	C	D	D
D	D	C	D	D	D	C	C	D	C	C
B	A	A	B	A	A	A	A	A	B	B
A	B	B	A	B	B	B	B	B	A	A
Question 17										
A	A	A	A	A	A	A	A	A	A	A
D	D	D	C	D	D	C	C	D	C	D
C	C	C	D	C	C	D	D	C	D	C
B	B	B	B	B	B	B	B	B	B	B
Question 18										
A	A	B	B	A	B	B	A	B	B	B
B	B	A	A	B	A	A	B	A	A	A
D	C	C	C	C	C	D	D	C	C	C
C	D	D	D	D	D	C	C	D	D	D
Question 19										
A	A	A	B	A	A	A	A	B	A	B
B	B	B	A	B	B	B	B	A	B	A
C	C	C	C	C	C	D	D	C	C	C
D	D	D	D	D	D	C	C	D	D	D
Question 20										
B	B	B	B	B	B	C	B	C	C	B
C	C	C	C	C	C	B	C	B	B	C
D	A	A	A	A	A	D	D	A	A	A

A	D	D	D	D	D	A	A	D	D	D
Question 21										
D	D	D	A	D	D	D	D	A	A	D
C	C	C	D	A	C	C	C	D	D	A
A	A	A	C	C	A	B	B	C	C	C
B	B	B	B	B	B	A	A	B	B	B
Question 22										
A	A	A	B	A	A	A	A	B	A	B
B	B	B	A	B	B	B	B	A	B	A
C	D	C	C	C	C	D	D	C	C	C
D	C	D	D	D	D	C	C	D	D	D
Question 23										
B	C	C	C	C	C	B	B	B	B	B
C	B	B	B	B	B	C	C	C	C	C
A	A	A	A	A	A	D	D	A	A	A
D	D	D	D	D	D	A	A	D	D	D
Question 24										
D	A	A	D	A	A	D	D	D	A	D
A	D	D	A	D	D	A	A	A	D	A
B	C	C	B	C	C	C	C	C	B	C
C	B	B	C	B	B	B	B	B	C	B