Establishing Local Norms for Two commercially available Numeracy Standardized Tests to identify Maltese Children with Mathematics Learning Difficulties

Esmeralda Zerafa
Chiswick House School, Malta
esmeraldazerafa@chs.edu.mt

Abstract: Mathematics Learning Difficulties (MLD) are of international and national concern. International research estimates that between four and seven percent of any population struggle with the learning of mathematics (Geary, 2004). Nonetheless, locally this field of research is still not adequately researched. Moreover, no numeracy assessment has been standardized with children in Malta. Consequently identifying children with MLD is based locally on using assessments which have been developed and standardized in other countries, in particular the U.K.. My doctorate research project aimed at finding effective strategies that help children to overcome their difficulties in Mathematics. The study was conducted with Grade 5 (9 to 10 years old) learners attending seven Catholic Church schools for boys. Six case studies were carried out with pupils attending the same school, who were selected to follow an intervention programme. The programme aimed at supporting learners with MLD to master the numeracy components that are fundamental for mathematics learning. This with the hope of finding effective strategies that would help learners struggling with mathematics to make the desired progress in the subject. This paper describes the process of sample selection. Three tests, which have been standardized in the U.K., were administered to a sample population of 352 boys out of the 704 boys attending Church schools for boys in Grade 5 and norms were established. The tests were then administered to all the boys attending Grade 5 at the school where I taught (50 pupils). The established local norms were then used to identify the boys with MLD who would participate in the intervention programme.

Keywords: Mathematics Learning Difficulties; numeracy asssessment; standardization of mathematics tests; sample selection

• • •

Introduction

Despite a similar prevalence rate to that of Reading Difficulties (RD) - an estimated four to seven percent (Geary, 2004) - the field of Mathematics Learning Difficulties (MLD) is still highly unexplored and under-researched in comparison to research on RD (Moeller, Fischer, Mag, Cress, & Nuerk, 2012). This is of concern since a number of studies (Bynner & Parsons, 2005; Poustie, 2000) have suggested that MLD may have great negative implications on a learner's school life and beyond. Following a longitudinal study carried out by Bynner and Parsons (2005), the researchers reported that MLD influence adults' life chances and therefore their quality of life. Bynner and Parsons highlight that adults who lack a basic grasp of numeracy skills are less likely to find a full-time job, to have access to an employer pension, and to be home owners. Moreover, they suggest that these adults are more likely to form part of a non-working household and to develop depression due to the lack of control over their lives. Bynner and Parsons also allude that difficulties in mathematics may have even higher negative impacts than RD. Their study concluded that adults having MLD were more likely to be unemployed than other adults who exhibited RD.

Interest from the international research community has recently increased (Moeller, Fischer, Mag, Cress, & Nuerk, 2012). Many studies focus on the neurobiological causes of MLD and, therefore, on the way the brain functions, and how this differs in children having MLD. Only a small number of studies have aimed at understanding what intervention strategies work with children having MLD. Hence, a wide knowledge lacuna still remains. There is, for instance, the need to understand what kind of intervention works with children having MLD so that we can make a difference in their learning trajectory. On a positive note, however, the international research indicates that children struggling with mathematics can make great progress if given the right form of intervention (Kaufmann, Handl, & Thöny, 2003). My doctorate study thus aimed at addressing the need to add to the existing international body of knowledge within this field by exploring what works with children having MLD or both MLD and RD through six case studies.

In Malta, awareness about MLD is still limited, and most schools still do not have an intervention programme to tackle MLD. In this scenario, my doctoral research is, to the best of my knowledge, the first of its kind. It aims at identifying effective strategies that support learners in mastering numeracy skills which, are crucial for mathematics learning in general, and more

specifically with learners in Malta. Moreover, it aims at developing a better understanding of the degree and nature of the MLD presented by learners with only MLD and those with both MLD and RD (MLDRD).

One important feature of conducting case studies is the selection of the appropriate participants. Selecting the right individuals ensures that they can serve as rich cases through which a phenomenon may be explored in depth. In my doctorate study, a main hurdle in this selection process was that no numeracy test had been standardized locally. Had I not decided to start by establishing local norms, I would have had to use the standardized scores found with a different population to identify the subjects of the local case studies. This might have rendered the tests invalid and unreliable, since the scores would not have pertained to children who live in Malta and who follow a similar educational system as the participants of the intervention programme. Local norms had to be established to ensure that the right participants could be selected. This paper will explain the process through which these norms were established. It will include the methods used for data collection as well as provide a summary of the results obtained.

Defining Terms, Identifying and Assessing for MLD

Research about MLD has been given increased importance in recent years; nonetheless, studies still refer to MLD using different terms that do not always refer to the same difficulties in mathematics learning. Different studies have made use of different terms. These include *Dyscalculia* (Chinn, 2012), *Developmental Dyscalculia* (Rubinsten & Sury, 2011), *Mathematical Learning Difficulty* (Hopkins & Egeberg, 2009), *Mathematics Disability* (Geary, 1993) and *Arithmetic Learning Disability* (Koontz & Berch, 1996). These terms seem to refer to a common difficulty: a difficulty with conceptualising and applying the essential number concepts and skills that are required in order to understand and actively participate in mathematical tasks (Geary & Hoard, 2001).

In my doctorate study each term was used purposely to indicate a specific construct. The term *Mathematics Learning Difficulties* (MLD) was used to refer to all the individuals underachieving in mathematics no matter what the underlying cause may be. Dowker (2005) suggests that terms like *Difficulties with Arithmetic, Mathematics and Numeracy* have been used generically to denote all "children or adults who struggle or fail to cope with some of the

aspects of arithmetic that are necessary or desirable for educational or practical purposes" (p.11). I made use of the term *Mathematics Learning Difficulties* because poor achievement in mathematics served as the fundamental criterion for the identification of the participants in the intervention programme. Knowledge of arithmetic is made up of a wide spectrum of skills (Dowker, 2005) and difficulties in this area are complex. It is well-known that learners with MLD are a heterogeneous group of individuals who may exhibit different difficulties which may stem from a variety of biological, genetic, social, and environmental causes (Bartelet, Ansari, Vaessen, & Blomert, 2014). Since the participants of my doctorate study would form part of this heterogeneous group of learners, it was deemed necessary to use this term to refer to this construct. On the other hand, the term *Dyscalculia* was used to refer to a specific learning difficulty in mathematics, and, therefore, only one type of MLD. This is substantiated by recent studies that illustrate that individuals with dyscalculia would probably exhibit an inability to perceive and understand the numerosity (the property of magnitude) of a set of objects (Geary, Hamson, & Hoard, 2009) and a difficulty in undertaking approximate number tasks (Piazza, Pinel, Le Bihan, & Dehaene, 2007). Moreover, dyscalculia may possibly be a genetically inherited condition (Ansari & Karmiloff-Smith, 2002).

Dowker (2005) suggests that there is no such thing as *arithmetical ability* but only *arithmetical abilities*. A learner normally underachieves in numeracy due to a weakness in one or more of the fundamental numeracy components that are the foundations for mathematics learning (Chinn, 2004; Dowker, 2004). Various studies highlight the key characteristics of individuals who struggle with mathematics, all of which seem to be related to number work. These include:

- Poor number sense (Bird, 2009; Emerson & Babtie, 2010);
- A delay in understanding some of the concepts of counting (Geary et al., 2000);
- A delay in using counting techniques for addition (Jordan & Montani, 1997);
- An over reliance on finger counting strategies (Ostad & Sorenson, 2007);
- A difficulty with sequencing (Emerson & Babtie, 2010);
- A deficit in various components of working memory (Roselli, Matute, Pinto, & Ardila, 2006).

Although an individual may have a deficit in either one or more of the areas mentioned above, recent studies have suggested that the main difference between learners with MLD and those with dyscalculia is that the latter learners usually have a poor sense of number and a deficit in interpreting numerosities (Emerson & Babtie, 2010; Halberda, Mazzocco, & Feigenson, 2008; Piazza et al., 2010).

The Diagnostic and Statistical Manual of Mental Disorders (DSM IV-TR) (American Psychiatric Association [APA], 2000) recommends three criteria for identifying whether a learner has Mathematics Learning Disabilities (MLD) or not. A learner may:

i.   Have lower performance in mathematics when compared to their attainment in other school subjects and general intelligence (IQ)[1];
ii. Score two or more *standard deviations* (SD) below the norm established by any mathematics standardised test;
iii. Not make expected improvement, even after appropriate classroom instruction.

The new Manual (APA, 2013) takes a different stance to the previous one (APA, 2000). It does not provide specific criteria for the identification of MLD but rather highlights criteria for identifying different Specific Learning Difficulties (SLD). As explained by Tannock (2012) in this new version of the DSM the definition provided is generic to SLD rather than MLD alone. This new definition is however more comprehensive as it does not focus mainly on IQ but sets new criteria for identification. These are:

- Having one of six symptoms[2] specified by the same manual which last at least 6 months;
- A discrepancy between actual age and achievement in the specific area;
- The learning difficulty becomes visible at the start of formal schooling;

---

[1] An Intelligent Quotient (IQ) is a score derived from one of many standarised tests made-up to assess intelligence. The IQ score defines one's intelligence in relation to the mean score of individuals on the same test.

[2] Four of these symptoms are related to literacy difficulties. The two symptoms identified in relation to number processing are: difficulty understanding number concepts, number facts and calculation and difficulty with mathematical reasoning.

- The learning difficulty is specific and not related to an intellectual disability.

I believe that having specific criteria for the identification of MLD is highly beneficial. This is because having specific criteria contributes to a stronger agreement as to which criteria are to be used to assess children having a profile of MLD. In my study, the term MLD is used to indicate learners who:

- Perform lower in mathematics when this attainment is compared to their attainment in other school subjects and general intelligence (IQ);
- Score two or more *standard deviations* (SD) below the norm established by any mathematics standardised test;
- Have difficulty understanding number concepts and number facts, struggle to perform calculations, and have problems with mathematical reasoning.

Identifying learners with MLD is not easy especially since multiple differences in definitions still exist resulting in a lack of instruments to assess for MLD. To date, I believe that the following assessment tools are currently valid tools for assessing for MLD: norm-referenced tests (also referred to as standardized tests), the use of direct observation, and the use of mathematical interviews. Using these modes of assessments ensures the proper detection of the characteristics of MLD, which I have highlighted. The use of formative assessments like the one proposed by Emerson and Babtie (2010), as well as screeners for Dyscalculia developed by Butterworth (2003), and by Trott and Beacham (2010), may also be of support in ensuring that a child is correctly assessed with a profile of MLD. In my study, The British Ability Scales (BASII) (Elliott et al., 1996) were used to test for IQ and be able to identify the first criterion. Two standardised numeracy tests (Gillham & Hesse, 2001; Chinn, 2012) were used to identify the second and third criteria. These were used in conjunction to interviews with parents and teachers as well as classroom observations to confirm specific difficulties being encountered in mathematics.

Standardized Tests

Standardized tests (STs), also referred to as Norm-referenced tests (NRTs), are the most popular means of assessing for MLD since they can show whether there is a gap between a learners' actual age and number age (age at which

the child is performing in numeracy), and, therefore, provide a clear indication of whether a learner truly has MLD. Results of STs can then be confirmed through other modes of assessment. Most STs focus on place-value, writing the numbers before and after a given number, the four operations (+, -, x, ÷), and continuing a sequence of numbers that follow a specific pattern (Butterworth, 2003; Chinn, 2012; Clausen-May et al., 2007; Emerson & Babtie, 2010; Gillham & Hesse, 2001). However, all STs will have different tasks, which are purposely graded to start from the easiest (the younger years) to more complex tasks (the older years). Every ST has a specific conversion grid that the assessor needs to use to convert the learner's raw score into a standardized score, a number age, a percentile, or a quotient. Thus, the main purpose of the ST is to assess the learner's mathematical achievement vis-à-vis their actual age and to identify a learner's number age (Shalev & Gross-Tsur, 2001). STs can be carried out on an individual basis or in groups.

Although STs have been used extensively for a variety of research projects, controversial issues still revolve around such tests (Higgins, 2009). A number of researchers (Gladwell, 2001; Phelps, 2005; Zwick, 2002) have debated the advantages and disadvantages of using standardized testing. STs are advantageous because they provide information about a learner's achievement that is more objective than that given through a teacher-created assessment. Numerous studies have indicated how teacher assessments may not be as accurate and valid as STs. Allal (2013), and, Wyatt-Smith and Klenowski (2013), indicate how teachers develop their own judgements about their pupils that then impinge on assessment. Moreover, Harlen (2004) suggests that teachers might have biases, such as those related to gender and special educational needs, which can impact assessment. The studies mentioned thus indicate that standardized testing may eliminate biases. As a result standardized tests are usually seen as more valid and reliable than teachers' observation (Marlow et al., 2014).

However, Miller (2003) has highlighted some disadvantages of standardized tests. These include that they create additional stress for teachers and learners, increase competition between students and schools, and are sometimes used to discriminate between groups of learners (Miller, 2003). Researchers (Gladwell, 2001; Phelps, 2005; Zwick, 2002) have also questioned the validity of test results resulting from these tests since they do not account for any factors that might impinge on test results such as the learner's mood

when doing the test. Notwithstanding, the arguments against standardized testing, I believe this method still remains an important and valid way of measuring a learner's achievement (Higgins, 2009), especially since they are objective. Moreover, the scores obtained from such tests allow the administrator of the test to compare the particular learner to others of his or her age. Keeping in mind that STs are not perfect, making use of a triangulation of assessment methods was a more accurate way of ensuring that the identification of the main participants was both valid and reliable. Since different STs test different spectra of mathematical content, the triangulation would allow me to confirm that the learner did have a profile of MLD and that their difficulties were not subject to the content being examined. The multiple assessments used as part of the triangulation process were: two standardized tests, summative assessments, as well as teachers' and parents' feedback about the child's achievement in mathematics. Results were also supported by other tests that would identify the characteristics of MLD. One of which was the BAS II (Elliott et al., 1996), which tested for IQ.

Cut-off Scores in STs

Every ST specifies a cut-off point, which indicates whether a learner has, in my case, MLD or not. Some also specify the degree of MLD as 'mild', 'moderate' or 'severe'. Cut-off points are test-dependent, so they have been cause for debate (Moeller et al., 2012). This has undoubtedly contributed to making it rather complex to develop a universal definition for MLD and dyscalculia. It has also augmented the difficulty of identifying the prevalence rate of MLD in the population. Different studies (Barahmand, 2008; Geary, 2010; Ramaa & Gowramaa, 2002) have used a varied range of cut-off scores for tests of similar difficulty, thus hindering researchers (Dirkset et al., 2008; Reigosa-Crespo et al., 2011) from agreeing on a universal prevalence rate. For example, a study conducted by Reigosa-Crespo et al. (2011) in Cuba made use of the 15th percentile as a cut-off point for their study. Their research indicated a prevalence rate of 3.4% for learners with MLD. On the other hand, Dirks et al. (2008), who carried out their study in the Netherlands, using a different standardized test, decided on a cut-off point of 25%, and this resulted in a prevalence rate of 10.3%.

In my doctorate study I made use of the wider construct of *MLD* as opposed to *dyscalculia*. This meant that I could include all those learners struggling with mathematics as long as they had at least an average IQ and the

characteristics highlighted earlier. The term MLD was taken to indicate all those pupils who fall below a cut-off point of approximately the 30th percentile. As a general rule, studies use this cut-off point to identify pupils who are underachieving in mathematics due to dissimilar potential causes without necessarily having a biological inherited weakness in mathematical cognition (Jordan, Kaplan, Olah, & Locuniak, 2006). This cut-off point allowed me to study a larger population of learners who are struggling with mathematics

Research Aims and Design

In my study, the use of a mixed methods approach was deemed to be very appropriate. Not having access to numeracy tests that have been standardized locally, I decided that finding local norms for the chosen standardized tests was the best way forward. This would allow me to select the participants for the intervention programme in a valid and reliable way. Following an evaluation of the processes involved in standardizing a test, I felt that it was sufficient to find norms for pupils having the same age as the eventual participants of the intervention programme (9 to 10 years old). Moreover, it was deemed suitable to find norms for pupils learning within the same educational setting as these eventual participants. Thus, I decided to standardize the test using a sample of pupils from all Church schools catering for boys.

Different STs were analysed and evaluated to find the most appropriate for the local context. Once three appropriate tests were selected these were pre-piloted with ten pupils attending the school where I taught, a Church school for boys. Following the pre-pilot study the choice of two tests out of the three piloted was concluded and a pilot study was conducted with an additional 15 pupils to ensure the suitability, reliability and validity of the tests to identify MLD. These three elements were checked for by comparing scores from the different tests as well as by carrying out observations during test administration. Teachers were also asked for their perception of the learners' abilities to compare these to the scores obtained on the tests. After analyzing the data obtained, two out of the three tests were considered suitable, reliable and valid. Following the pilot study, the standardization exercise commenced. The first step of this process was to administer the tests to a representative sample population. It was thus important to determine the sample size so as to understand how much time and resources would be

needed for the actual collection of data. As suggested by Gogtay (2010), "Sample size calculations must take into account all available data, funding, support facilities, and ethics of subjecting patients to research" (p. 517). It primarily needs to consider what type of data the research is dealing with, whether quantitative or categorical. In this case, the data was of a quantitative type. An online sample size calculator (Creative Research Systems, 2012) indicated that if scores were collected for half the population of boys in Grade 5 attending Church schools, i.e. 352 out of the 704 boys, the statistical power of the results would be sufficient in making the result valid and reliable. This sample population was calculated with a confidence level of 95% and a confidence interval of 1.2. Administering the tests to half the population took into consideration a safety margin and dropout rate.

When the sample size was determined, access was sought from the relevant entities including the Maltese Episcopal Curia and the Heads of each individual school in which the tests were to be administered. Some schools had acquired consent from parents prior to the scholastic year to carry out assessments as deemed fit but others had not. In the latter case, consent was also acquired from parents. All children were also asked for consent before the test was administered and were free to withdraw if they wanted to. A number of precautions to maintain the tests' validity and reliability were taken whilst administering the STs. These included:

  i.   Administering all the tests myself;
 ii.   Introducing myself and informing the learners what the test was going to be used for;
iii.   Reassuring the learners that they needed to try their best ensuring that they did not feel stressed by the test itself.
 iv.   Administering the tests to the class as a whole.

Selecting an Appropriate Standardized Test

After evaluating different tests, I decided that two tests were most appropriate: the Basic Number Screening Test (BNST) (Gillham & Hesse, 2001) and Chinn's Number Tests (CNT) (2012). A third test, Progress in Mathematics (PIM) (Clausen-May et al., 2009), was also identified as being appropriate because it was specifically designed to use with learners in Grade 5 and had also been more recently published than the BNST. The decision of evaluating all three tests before defining the two to be used was based on two

main factors. Primarily, all three assessments were in line with our curriculum. The exercises set are tasks that focus on the same algebra and number work determined by our curriculum. Secondly, the assessments focused mostly on number operations and calculation rather than measures, data handling, shape and space – this was important since the vast majority of children with MLD display difficulties with the former areas of mathematics rather than the latter (Dowker, 2005; Emerson & Babtie, 2010).

CNT is a standardized test that can be used with both children and adults and has been standardized with different age groups in the UK. CNT does not include any written instructions but the learners merely need to work out the computations given. This feature is deemed important in the actual assessment of MLD since valid assessment should exclude other difficulties that may diagnose a learner with MLD incorrectly. This assessment has multiple parts but its main one is an assessment of the four operations (+, -, x and ÷) involving whole numbers. The pupils are first asked to work out as many addition sums as possible in one minute (maximum of 36). This exercise is repeated with subtraction sums and the pupils are given another minute to work out as many as they can out of 36 subtraction sums given. A sheet of 33 multiplication sums and another of 33 division sums follow these two exercises. The pupils are given two minutes to complete each of the multiplication and division sheets. This assessment is then followed by a 15-minute assessment which involves different types of computations including not only the four main operations with whole numbers but also simple fractions, percentages, conversions of time and length, and working with decimals. The computations vary in difficulty and are graded starting with easier tasks which become more complex in nature. All assessments were carried out with each class as a whole, reducing the variables linked to administering them on a one-to-one basis. It also made the exercise feasible, as it would have otherwise been too time consuming.

CNT has other parts of it that assess for mathematics anxiety and other basic mathematics skills. It also allows the identification of learning styles related to mathematics learning. However, these components are to be administered on a one-to-one basis and could not be standardized. As a result, due to the large numbers involved, these were not used in the process of collecting norms. These parts were however administered later to the six participants chosen for the intervention programme to gain a deeper insight into the learners' characteristics and how they learn mathematics.

The BNST was chosen because it has no written instructions; assesses a wide range of numeracy components, including the addition, subtraction, multiplication and division of whole numbers as well as fractions; and, only takes roughly 25 minutes to complete. The test was developed by Gillham and Hesse (2001) through Hodder Education and has also been standardized in the UK. The test is suitable for learners aged 7 to 12. When the test is administered the assessor has to read out the instructions for one task and the children have to complete it. Each instruction is read out twice. The instructions to the tasks are in English. Due to this, when this test was administered during the pre-pilot study, I decided to translate all the instructions to Maltese as I did not want the children's understanding of the English language to influence the score obtained. Hence, when administering the tests during the pilot study and during the actual study, I read out the instructions in both English and Maltese prior to every task. Translating the instructions involved multiple steps. These were:

i. Translating the instructions myself;
ii. Having the translation checked by a mathematics educator and a linguist;
iii. Making any changes required according to the suggestions given by the reviewers of the instructions;
iv. Having a professional translator do a back translation of the content to ensure that words and phrases were correctly interpreted. Through this process, changes to be made at word and sentence level were indicated. These changes ensured that the same meaning was attributed to the instructions in Maltese as in the English version.

PIM has different tasks for learners at different levels. For the purpose of this study I used the PIM 5, a test suitable for Grade 5 learners. All questions set are in English and it contains written questions that the learner needs to solve. The assessor may read the instructions in order to eliminate the reading variable. It takes roughly 45 minutes to complete the test. The content of the test is similar to that found in the BNST and CNT, thus focusing mostly on number work too.

As the triangulation of results would render my findings more robust, I decided that it was best to use two of the tests rather than only one to be able to compare results and ensure that the learners identified as having MLD truly did. Although the tests were not exactly the same, they tested similar

numeracy components. A pre-pilot study was carried out to determine which of the three tests identified as appropriate was most suitable for detecting MLD in local school children.

The Pre-Pilot Study

In the pre-pilot phase, ten pupils were chosen who were then in Grade 5 at the school where I taught. The Head of School granted access and consent to administer the test to the pupils was gained from both parents and pupils. Class teachers were asked to identify learners with a range of abilities for this exercise: two were low achievers; six were average achievers, and another two were high achievers in mathematics. This was crucial since I was interested in comparing how different pupils would perform in the three assessments. All ten pupils sat for the BNST. Out of the same ten pupils, five sat for CNT and the other five sat for PIM. Both subgroups were composed of one low achiever, three average achievers and one high achiever. One of the reasons why I did not give all three tests to all ten pupils was that the learners would miss too much learning time from class to complete all tests. Another was that they would have probably become too bored doing all three and the results would not have remained valid.

The data from the pre-pilot study was analysed by looking at the scores obtained by each individual pupil in the two tests that were administered to him. Using this methodology allowed me to compare the scores obtained in the two tests in order to highlight commonalities and discrepancies in performance. It was assumed that this would help me to determine the validity of a test in identifying MLD; if the scores obtained in both tests were comparable, the detection of MLD would be more accurate. In Table 1 I have illustrated the percentile scores obtained on the BNST and CNT assessment as well as the teacher's perception of the each learner's mathematical abilities. The CNT only offers percentile scores when the learner scores below the 30th percentile, if the pupil scores higher than the 30th percentile, a comment such as 'average' or 'above average' is provided to describe the pupil's achievement.

Pupil Code	BNST Percentile Score	CNT - 15-mins. Percentile Score	CNT - Addition, Subtraction, Multiplcation, and Division Percentile Score	Teacher's Perception of Child's Maths Ability
PP1	5th	Below 5th	Addition: 10th Subtraction: 10th Multiplication: 5th Division: 10th	Below Average
PP2	40th	Between 10th and 5th	Addition: Average Subtraction: Average Multiplication: 10th Division: Above Average	Average
PP3	80th	25th	Addition: Average Subtraction: Average Multiplication: 10th Divi. Score: Average	Average
PP4	90th+	80th – 75th	Addition: 10th to 5th Subtraction: 25th - 10th Multiplication: 5th Division: Above Average	Above Average
PP5	90th+	50th – 40th	Addition: Above Average Subtraction:Above Average Multiplication: Average-25th Division: Above Average	Average

Table 1: A comparison of scores obtained in the BNST versus those obtained in CNT's 15-minutes assessment and that for the four operations as well as the teacher's perception of learner's mathematical abilities

The scores for the BNST and CNT were compared by looking at each of the percentile scores achieved by each individual learner on both tests. During this comparison I took note of whether the scores complemented each other. When the scores were similar it meant that both tests had placed the child within the same category of learners (i.e., 'average', 'below average', etc.). Discrepancies in scores, on the other hand, meant that the different tests were placing the same learner in two different categories. The scores obtained in both tests were also compared to the class teacher's perception of the child's mathematical abilities. A similar exercise was also carried out with the other five learners sitting for the BNST and PIM. This comparison is presented in Table 2.

Pupil Code	BNST Percentile Score	PIM Standardised Score	Comment Retrieved from PIM Manual	Teacher's Perception of Child's Maths Abilities
PP6	90th+	113	Above Average	Above Average
PP7	90th+	101	Average	Average
PP8	60th	106	Average	Average
PP9	70th	87	Below Average	Lower Average
PP10	20th	69	Very Low	Below Average

Table 2: A comparison between scores obtained in the BNST, PIM and the teacher's perception of learners' mathematical abilities

After the analysis of all the data collected from the pre-pilot study, some discrepancies in the results obtained in the different tests were evident (see, for instance, PP2, PP3, PP5 and PP9). One possible reason for the discrepancy between the scores obtained in the BNST and CNT may have been that CNT covers topics that are higher in level of difficulty since some of the computations require more complex calculation skills. Another reason may be that CNT is timed, and, therefore, the pupils' speed of working out through the tasks would have impinged on the score obtained. On the other hand, when comparing results from the BNST and PIM, discrepancies in the scores were possibly due to the learner not being able to understand the instructions since the latter involved written questions indicating what the learner had to do to complete the task. Although some discrepancies in the scores obtained by the learners in all three test were noted, some similarities in the outcomes were also evident. For example, the scores obtained agreed that PP1 and PP10 had severe MLD and that PP2 had mild-to-moderate MLD. The conclusions from the tests also supported the teachers' perception of the pupils' abilities in mathematics.

It was decided that all three tests could be considered as reliable since most of the results between the tests were coherent. However, the BNST and CNT were the assessment tools deemed most appropriate. The fact that PIM included a lot of written instructions may have resulted in an invalid picture

of the learners' abilities since a pupil may have been assessed with a profile of MLD due to the nature of his/her reading difficulties. Moreover, it took the children a long time to complete PIM: approximately between 45 minutes and an hour, which contrasted with the 20 to 25 minutes employed to complete each of the BNST and CNT. Due to this, I could observe that some pupils got bored and began to guess answers to solve the questions in the PIM. This was not observed for the other two tests, indicating that using PIM, in comparison to the BNST and CNT, might have increased the risk of obtaining invalid findings.

The Pilot Study

During the pilot study CNT and the BNST were administered in this respective order. Only the BNST was translated to Maltese because CNT has no instructions, and, therefore, no translations were needed. During this piloting phase, it was deemed necessary to administer both tests. This was done to be able to determine whether the back translation for the BNST was fine and whether there was anything else that could be done differently during the actual data collection process.

Pupil Code	BNST Score	CNT - 15-min. Assessment Scores	CNT – Addition (A), Subtraction (S), Multiplication (M) and Division (D) Scores	Teacher's Perception of Child's Maths Abilities
P1	90th percentile	80th percentile	A: Above Average S: Above Average M: Above Average D: Above Average	Above Average
P2	40th percentile	20th percentile	A: 45th percentile S: 25th percentile M: 35th percentile D: Average	Low
P3	50th percentile	12th percentile	A: Above Average S: Above Average M: Above Average D: Above Average	Low
P4	60th percentile	60th percentile	A: Average S: Average M: 40th percentile D: Above Average	Average
P5	80th percentile	77.5th	A: Above Average	Average

P6	90th percentile	75th percentile	A: Above Average S: Above Average M: Above Average D: 25th percentile	Average
P7	80th percentile	75th percentile	A: Above Average S: Above Average M: Above Average D: 40th percentile	Average
P8	60th percentile	20th percentile	A: 25th percentile S: 35th percentile M: 40th percentile D: Average	Average
P9	40th percentile	20th percentile	A: 20th percentile S: Average M: Above Average D: Above Average	Low
P10	90th percentile	90th percentile	A: Above Average S: Above Average M: Above Average D: Above Average	Above Average
P11	90th percentile	80th percentile	A: Above Average S: Above Average M: Above Average D: Above Average	Above Average
P12	80th percentile	75th percentile	A: Above Average S: Above Average M: Above Average D: Above Average	Above Average
P13	40th percentile	35th percentile	A: 25th percentile S: 30th percentile M: 45th percentile D: 45th percentile	Low
P14	60th percentile	50th percentile	A: 35th percentile S: Average M: 45th percentile D: 35th percentile	Average
P15	70th percentile	55th percentile	A: 35th percentile S: 45th percentile M: 20th percentile D: Above Average	Average

Table 3: Scores obtained in the pilot study by the 15 participants

Fifteen Grade 5 children (aged 9 to 10) were chosen randomly from the school where I taught. Random selection was used to ensure that the learners had different mathematical abilities. The Grade level teachers were asked to nominate four children who were low-achievers, seven average-achievers, and four high-achievers. The results obtained by these learners can be seen in Table 3.

The results obtained were generally comparable, since most of the pupils obtained similar results in both tests. This indicated that the two tests could be used in tandem so as to collect more valid and reliable data. Pupils like P1, P10, P11 and P12 clearly scored an above average score in both tests indicating that they did not have MLD. On the other hand, pupils like P2 and P13 seemed to be struggling with mathematics since all their tests scores indicate this. When pupils' scores were not coherent, such as in the case of P3, possible reasons for this were looked into. It seemed that P3 was able to successfully complete the simple addition, subtraction, multiplication and division sums but then found difficulty in both the other tests, that is, the BNST and the 15-minute test component of CNT. The latter contains more complex mathematical tasks including work with fractions and percentages. I thus checked with the class teacher to identify whether this pupil was struggling with mathematics and the teacher said he was. Anyway, since the pupil's difficulty in mathematics was detected by two of the tests, it was concluded that the test results were valid and reliable. Some other discrepancies with the scores obtained by the children were also noted. The scores obtained in the two numeracy tests by P8, P9 and P14 were slightly different since all three performed better in the BNST as opposed to CNT. Nonetheless, the discrepancy was minor and still placed the learner within the same category of achievement (i.e. whether low achieving, average or high achieving).

Following this piloting phase, I decided to change the order in which the tests were administered. Whilst carrying out the pilot study I realized that the children enjoyed doing CNT more than the BNST. This was possibly due to the fact that CNT test is timed. The children seemed to enjoy this characteristic of CNT as they took it up as a challenge to complete as many sums as possible in the time given. Thus, when administering the tests with the larger sample, the BNST was administered first. This was done with the hope of maintaining the children's motivation throughout both tests so that they would not guess answers due to boredom, and, thus, invalidate results.

The Data Collection Stage

Once access was granted by the relevant entities, appointments were set to administer the tests in the seven schools. Various validity and reliability recommendations were taken into account to maintain a high level of these in the data collection process (Cohen et al., 2007; Lincoln & Guba, 1985):

i. All the tests were administered by myself to all learners ensuring that consistency was maintained in reading speed or fluency. In this way, additional variables that could have been problematic had the tests been administered by multiple persons were avoided;

ii. As for the pre-pilot and the pilot study, in the actual collection of data, the children were asked to cover their work or to separate their desks so that copying was evaded;

iii. As much as possible I sought to administer the tests to the pupils during the same period of time. In fact, all data was collected between November and December 2012, so that the pupils would have covered approximately the same topics in class, since as confirmed by the schools, all had started using the same textbook and had covered roughly the same topics. For the same reason, this period was considered suitable a year later, when assessing participants for the case studies. This same period was also then maintained the following year when assessing the participants for the case studies;

iv. When possible, tests were also administered at the same period of time during the day, i.e. the morning. This choice was based on the fact that children are usually less restless at this time of day;

v. All tests were administered in the children's own classroom with the presence of the teacher to ensure that the children felt secure and safe in a familiar environment.

Analysis of Data

All the tests were scored by myself. The raw scores were entered on SPSS and a z-score (standardized score) was computed for each raw score. These z-scores were then saved as variables and were later used to find the norms. The quotient was then calculated through MS Excel by using the formula 'z-score * 15 + 100'. Finally, percentiles were also smoothed. The cut-off points used were as for 99 equal groups. Since the data collected was of ordinal type, it was not assumed that the difference between one score and another

was equivalent to the interval between any other two percentiles. Therefore, for example, the difference in score between the 98th and 99th percentile could have been much larger than that between the 50th and 55th score. The tests used to analyse the data were non-parametric tests, which correlate with the type of data collected since the data had independent variables. Through this data analysis process, the median, quartiles and percentiles in multiples of 5 were worked out (5th, 10th, 15th, 20th, 25th etc.). The crucial percentile and related score needed for the identification of the main participants of the main study was the 30th percentile scores since the cut-off point I used for this research was the 30th percentile. The raw scores obtained as norms for these three percentile scores are illustrated in Table 4.

Assessment	30th Percentile Score
CNT - Addition	18 and below
CNT - Subtraction	16 and below
CNT - Multiplication	20 and below
CNT - Division	16 and below
CNT - 15-minute assessment	15 and below
Basic Number Screening Test	22 and below

Table 4: Scores extrapolated for the 30th percentile following the collection and statistical analysis of the data collected in this study

The local norms found for the specific population of Grade 5 boys attending Church schools for boys were compared to the norms achieved in the U.K. for all tests. The latter norms are ones that have been established through a sample population of the whole population of Grade 5 children in the U.K.. It was considered interesting to explore how the cohort for whom the local norms were found, actually compared to the general cohort of Grade 5 pupils in the U.K.. In Table 5, I present the local and U.K. norms for the 30th percentile.

When comparing the local established norms to the ones found in the U.K. for Grade 5 pupils, the following observations were made. Primarily, the U.K. and local scores for Chinn's (2012) assessment were very similar. In fact, the U.K. and local norms for the addition and subtraction components were identical. Moreover, the local norms for the multiplication and division components, as well as those for the 15-minute assessment, were only slightly higher than those found in the U.K.. On the other hand, an important finding was that the local norms for the BNST are higher than the U.K. norms. This

was in line with the findings from the pre-pilot and the pilot study in which some pupils did well in the BNST, but not so well in CNT. This indicates that the local population for whom the norms were found - boys attending Church schools - performed generally better in the mathematics components assessed in this test than the population with whom this test was standardized with in the U.K.. Despite this result, one must highlight that had the test been administered to a wider population in Malta, including girls and other sectors of the local education system, the findings may have been different and the difference not as accentuated.

Assessment	30th Percentile Local Score	30th Percentile UK Score
CNT - Addition	18 and below	18 and below
CNT - Subtraction	16 and below	16 and below
CNT - Multiplication	20 and below	19 and below
CNT - Division	16 and below	12 and below
CNT - 15-minute assessment	15 and below	13 and below
Basic Number Screening Test	22 and below	14 and below

Table 5: 30th percentile scores obtained by Maltese Grade 5 boys attending Church schools compared with 30th percentile scores obtained by the whole population of U.K. Grade 5 pupils

Another interesting observation was that the discrepancy between the norms achieved for the BNST and CNT test shows that, in general, the content covered in CNT, although testing similar numeracy components, was found to be more difficult than that presented in the BNST. This finding reflects the U.K. norms for both tests too, since this same discrepancy is also evident when these are compared.

Conclusions, Limitations and Recommendations for Further Research

Through this phase of my doctorate study I found norms for numeracy assessments for one group of learners – boys in Grade 5 (ages 10 to 11) Church schools. In this paper, the local norms collected were discussed and were compared to those collected in the U.K. The local norms established during this phase of my doctorate study were crucial to my intervention programme, as they allowed me to identify in a valid manner the six main participants for the qualitative part of the study. This qualitative part

consisted of six case studies. The scores obtained by the cohort at the school where I taught were compared to the established local norms. Pupils who achieved scores that were equal to or below the 30th percentile were then assessed using further tests, for example the BAS which assesses for IQ, to ensure that they had the characteristics identified in learners with MLD. These participants were also confirmed by asking for the teacher's feedback about the children's achievement in mathematics and by looking at their previous examination paper marks (those taken at the end of Grade 4). Indeed, four out of the six pupils had failed their mathematics examination. The other two had just managed to get a pass mark. Thus, having been identified with a number of criteria for MLD, these learners were asked to participate in the intervention programme.

The norm collection process was carried out for only one specific group of learners (i.e., Grade 5 boys attending Church schools for boys). Due to this, norms for other groups of learners, such as those in other levels, in other educational settings and girls, were not found. This is a limitation of this part of my study and thus, there is a need for the process of establishing norms to be replicated for different groups of learners. Educators urgently need to acquire assessment tools that accurately identify learners struggling with mathematics. This need arises from an increased awareness about MLD and the impact they might have on an individual's life. Difficulties with mathematics can persist throughout adult hood reducing life opportunities such as employment (Bynner & Parsons, 2005). Hence, these tools are essential for the early identification of mathematics learning difficulties.

References

Allal, L. (2013) Teachers' professional judgement in assessment: a cognitive act and a socially situated practice. *Assessment In Education: Principles, Policy & Practice*, *20*(1), 20-34.

American Psychiatric Association. (APA) (2000) *DSM IV-TR: Diagnostic and statistical manual of mental disorder.* Washington, D.C.: American Psychiatric Association.

American Psychiatric Association. (2013) *Diagnostic and statistical manual of mental disorders* (5th Edn.). Washington, DC: Author.

Ansari, D., & Karmiloff-Smith, A. (2002) Atypical trajectories of number development: a neuroconstructivist perspective. *Trends In Cognitive Sciences, 6*(12), 511-516

Barahmand, U. (2008) Arithmetic disabilities: training in attention and memory enhances arithmetic ability. *Research Journal of Biological Sciences, 3*(11), 1305–1312. Retrieved http://docsdrive.com/pdfs/medwelljournals/rjbsci/2008/1305-1312.pdf, 12th August, 2016.

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014) Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities, 35*(3), 657-670.

Bird, R. (2009) *Overcoming difficulties with number: supporting dyscalculia and students who struggle with maths.* London: Sage Publications.

Butterworth, B. (2003) *Dyscalculia screener: Highlighting children with specific learning difficulties in mathematics.* London: NFER-Nelson.

Bynner, J. & Parsons, S. (2005*) Does Numeracy Matter More?* Research carried out for the National research and development center for adult literacy and numeracy. Retrieved http://eprints.ioe.ac.uk/4758/1/parsons2006does.pdf, 25th February, 2015.

Chinn, S. J. (2004) *The trouble with maths: a practical guide to helping learners with numeracy difficulties.* London: Routledge Falmer.

Chinn, S. J. (2012) *More trouble with maths: a complete guide to identifying and diagnosing mathematical difficulties.* London: Routledge.

Clausen-May, T., Vappula, H., Ruddock, G. & NfER (2009) *Progress in maths* (PIM). London: GL Assessment.

Cohen, L., Manion, L., & Morrison, K. (2000) *Research methods in education* (5th Ed..). London: Routledge.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education.* (6th Ed.). London: Routledge Falmer.

Creative Research Systems. (2012) *Sample Size Calculator.* Retrieved https://www.surveysystem.com/sscalc.htm, 3rd August, 2013.

Dirks, E., Spyer, G., van Lieshout, E., & de Sonneville, L. (2008) Prevalence of Combined Reading and Arithmetic Disabilities. *Journal Of Learning Disabilities, 41*(5), 460-473.

Dowker, A. (2005) *Individual differences in arithmetic*. Hove [U.K.]: Psychology Press.

Dowker, A. (2004) *What works for children with mathematics difficulties?* (RR554) London: DfES. Retrieved from http://www.catchup.org/LinkClick.aspx?fileticket=59GXj0uNY1A%3d&tabid=105, 30th August, 2012.

Elliott, C. D., Smith, P. and McCulloch, K. (1996) British Ability Scales Second Edition (BAS II). Administration and Scoring Manual. London: Nelson.

Emerson, J. & Babtie, P. (2010) *The dyscalculia assessment.* London: Continuum.

Geary, D. C. (2004) Mathematics and learning disabilities. *Journal of Learning Disabilities, 37, 4-15*.

Geary, D. C. (1993) Mathematical disabilities: cognition, neuropsychological and genetic components. *Psychological Bulletin, 114*, 345-362.

Geary, D. C. (2010) Mathematical disabilities: Reflections on cognitive, neuropsychological, and genetic components. *Learning and Individual Differences, 20*(2), 130-133.

Geary, D. C., & Hoard, M. K. (2001) Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in pupils with learning disability. *Journal of Experimental Pupil Psychology, 54*, 372-391.

Geary, D. C., Hamson, C., & Hoard, M. (2000) Numerical and Arithmetical Cognition: A Longitudinal Study of Process and Concept Deficits in Children with Learning Disability. *Journal Of Experimental Child Psychology, 77*(3), 236-263.

Gillham, B. & Hesse, K. (2001) *Basic number screening test manual.* London: Hodder Education.

Gladwell, M. (2001) Examined life. *New Yorker*, 17th December, pp. 86-92.

Gogtay, N. J. (2010) Principles of sample size calculation. *Indian Journal of Ophtalmology, 58*(6), 517-518.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008) Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665-668.

Harlen, W. (2004) *A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes* (EPPI-Centre Review). Research Evidence in Education Library, Issue 3. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Retrieved from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=116, 16th June, 2016.

Higgins, M. (2009) Standardised Tests: Wristwatch or Dipstick?. *Research In Education*, *81*(1), 1-11.

Hopkins, S., & Egeberg, H. (2009) Retrieval of simple addition facts: complexities involved in addressing a commonly identified mathematical learning difficulty. *Journal of Learning Disabilities, 42*(3), 215-229.

Jordan, N. C., Kaplan, D., Olah, L N., & Locuniak, M. N. (2006) Number sense growth in kindergarten: a longitudinal investigation of children at risk for math difficulties. *Child Development, 77*, 153-175.

Jordan, N., & Montani, T. (1997) Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *Journal Of Learning Disabilities*, *30*, 624-634.

Kaufmann, L., Handl, P., & Thöny, B. (2003) Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: A pilot study. *Journal of Learning Disabilities, 36*(6), 564-573.

Koontz, K. L., & Berch, D. B. (1996) Identifying simple numerical stimuli: processing inefficiences exhibited by arithmetic learning disabled pupils. *Mathematical Cognition, 2*(1) 1-23.

Lincoln, Y. S. & Guba, E. G. (1985) *Naturalistic Inquiry.* Newbury Park, CA: Sage Publications.

Marlow, R., Norwich, B., Ukoumunne, O., Hansford, L., Sharkey, S., & Ford, T. (2014) A comparison of teacher assessment (APP) with standardised tests in

primary literacy and numeracy (WIAT-II). *Assessment In Education: Principles, Policy & Practice*, *21*(4), 412-426.

Miller, M. (2003) *The two per cent solution.* New York (NY): Harcourt.

Moeller, K., Fischer, Mag, U., Cress, U., & Nuerk, H. (2012) Diagnostics and Intervention in Developmental Dyscalculia: current issues and novel perspectives. In Z. Breznitz (Ed.) *Reading, Writing, Mathematics and the Developing Brain: listening to many voices* (1st Ed.) (pp. 233-275). Dordrecht: Springer.

Ostad, S. A., & Sorenson, P. M. (2007) Private speech and strategy-use patterns: bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *Journal of Learning Disabilities, 40*(1), 2- 14.

Phelps, R. (2003) *The war on standardized testing: kill the messenger.* New Brunswick NJ: Transaction Publishers.

Piazza, M., Facoetti, A., Trussardi, A.N, Berteletti, I., Conte, S., & Lucangeli, D., Dehaene, S., & Zorzi, M. (2010) Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33-41.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007) A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. Neuron, *53*(2), 293-305.

Poustie, J. (2000) *Mathematics solutions: an introduction to dyscalculia: part A – how to identify, assess and manage specific learning difficulties in mathematics.* Taunton: Next Generation.

Roselli, M., Matute, E., Pinto, N., & Ardila, A. (2006) Memory abilities in children with subtypes of dyscalculia. *Developmental Neuropsychology 30*(3), 801-818.

Rubinsten, O., & Henik, A. (2009) Developmental dyscalculia: heterogenity might not mean different mechanisms. *Trends in Cognitive Sciences*, *13*(2), 92-99.

Ramaa, S. & Gowramma, I. P. (2002) A systematic procedure for identifying and classifying children with dyscalculia among primary school children in India. *Dyslexia, 8*, 67–85.

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Santos, E., & Torres, P. (2011) Basic numerical capacities and prevalence of developmental dyscalculia: the havana survey. *Developmental Psychology, 48*(1), 123-135.

Rubinsten, O., & Sury, D. (2011) Processing Ordinality and Quantity: The Case of Developmental Dyscalculia. *Plos ONE*, *6*(9), e24079.

Shalev, R., & Gross-Tsur, V. (2001) Developmental dyscalculia. *Pediatric Neurology*, *24* (5), 337-342.

Tannock, R. (2012) Rethinking ADHD and LD in DSM-5: Proposed Changes in Diagnostic Criteria.*Journal Of Learning Disabilities*, *46* (1), 5-25.

Trott, C & Beacham, N. (2010) *DysCalculiUM: a first-line screening tool for dyscalculia.* Cambridge: Iansyst Ltd.

Wyatt-Smith, C. & Klenowski, V. (2013) Explicit, latent and meta-criteria: types of criteria at play in professional judgement practice. *Assessment In Education: Principles, Policy & Practice*, 20(1), 35-52.

Zwick, R. (2002) *Fair game: the use of standardised admissions tests in higher education.* New York: RoutledgeFalmer.