Teaching and Learning Practices in Secondary Mathematics: measuring teaching from teachers' and students' perspective

Maria Pampaka, Lawrence Wo, Afroditi Kalambouka, Sophina Qasim, David Swanson

Abstract

This paper presents initial findings from an ongoing ESRC funded study of teaching and learning secondary mathematics in UK (www.teleprism.com). The substantive aim of this study is to understand (i) how learners' dispositions to study mathematics develop through secondary school, (ii) how mathematics pedagogies vary across different situations and contexts and (iii) how these pedagogies influence learning outcomes. The research question we seek to answer in this paper regards measuring teaching practices. In particular, we set out to measure teaching practices from students' and teachers' point of view.

For this analysis we draw on the first data point (out of three) of our longitudinal survey of students in Year 7 to 11 (N=13,000+) and their mathematics teachers, which took place during the previous academic year (October to December 2011). This involved a questionnaire about students' attitudes to mathematics, confidence at various mathematical topics, future aspirations, and their perceptions of the teaching they encounter. The latter was also captured through a teacher survey administered to their mathematics teachers and will serve as the focus of the paper. Validation is performed within the Rasch measurement framework, seeking validity evidence through fit statistics and Differential Item Functioning (for measure invariance across year groups). Once the measures' validity is established we investigate the degree of agreement between teachers' and students' perceptions of teaching and use these measures in further statistical modelling with students' dispositions and attainment measures.

Acknowledgement:
ESRC Funded under the First Grant Scheme (2009-10 call): RES-061-25-0538
Title: Mathematics teaching and learning in secondary schools: the impact of pedagogical practices on important learning outcomes.

1. Introduction – Theoretical Background

In our previous work, and also a keynote symposium presentation in this conference (Pampaka, Pepin & Sikko, Keynote Symposium 3) we demonstrated the damaging effect of ‘transmissionist’ pedagogies in school/college (pre-university) students’ mathematics dispositions. Drawing on these findings, and others from our previous TransMaths[1] projects, this new study, entitled ‘Teaching and Learning Practices in Secondary Mathematics’ (TeLePriSM), aims to map secondary students’ learning outcomes, attitudes and choices regarding mathematics, together with the teaching they are exposed to.

The Teleprism study is substantially driven by the STEM (Science, Technology, Engineering and Mathematics) agenda and the significance of secondary school mathematics for this: it builds on the work of the TransMaths team, and attempts to fill in the knowledge gap (i) in how students’ dispositions develop, and (ii) in how pedagogy and other school variables influence this development, and how these influence key outcomes and decisions at age 16.

The study is grounded in a varied and rich theoretical base, which can only be summarised here, around some focal questions.

The first question is ***‘Why Mathematics?’*** Our study’s focus is on mathematics because of its importance to students’ access to STEM subjects, and hence to their educational and socioeconomic life opportunities (OFSTED, 2006; Roberts, 2002; Smith, 2004; Wolf, 2002). In a recent report ACME (ACME, 2009) recognises this important issue and advocates ‘tackling the perceptions of mathematics” as a particularly important issue in the current economic climate, placing emphasis on the importance of mathematics as a “powerful analytical tool”, with inherent “pervasiveness” and a “key workforce skill”. They note: “urgent attention must be focused on the impact the move to two-tier GCSE mathematics is having on KS4 teaching and progression to Level 3”. From the schools’ and teachers’ perspective, the topic is also very timely: schools are increasingly finding that raising outcomes in mathematics - after some short term 'fixes' have done their work on test scores - requires sustained attention to the *quality of pedagogy* and to *student engagement* throughout the secondary experience. This leads us to the next questions:

‘Why study student engagement and dispositions?’ The study of students’ dispositions is very important because this may reveal key influences on their choices and decision-making and hence future engagement with STEM (Archer, Halsall, Hollingworth, & Mendick, 2005; Pustjens, Van de gaer, Van Damme, & Onghena, 2004; Wolf, 2000) (Allexsaht-Snider & Hart, 2001). Previous studies had also identified a plethora of socio-cultural factors which are significant in shaping

[1] TransMaths website: www.transmaths.org

students' dispositions and choice-making in education in general, and in STEM subjects and mathematics in particular: class, gender, nationality, ethnicity, parental and peer cultures are just the beginning of the list (Cao, Bishop, & Forgasz, 2007; Mendick, 2005; O'Brien, Martinez-Pons, & Kopala, 1999; van Langen & Dekkers, 2005). Students' affective dispositions (e.g. self-efficacy) may also be critical to their choices and need to be included in modelling learning outcomes alongside traditional indicators such as grades (Bandura, 1978; Bandura & Locke, 2003; Bong, 2001, 2004; Marat, 2005).

And finally, "***why bother with the quality of pedagogy in maths?***" The general argument for the importance of the quality of mathematics teaching is well documented (Askew, Brown, Rhodes, Johnson, & Wiliam, 1997; de Jong, Westerhof, & Kruiter, 2004; Driessen & Sleegers, 2000; McCaffrey, et al., 2001; Schuh, 2004; Swan, 2000; Swanson & Stevenson, 2002; Vermunt & Verloop, 1999). In our previous work we have also shown how some institutional and pedagogic practices can encourage reduction of the learning of mathematics to the instrumental level only, thus limiting future educational opportunities (Pampaka, et al., 2012; Wake & Pampaka, 2008): many have argued that formative assessment and more dialogical pedagogies are required for conceptual, metacognitive, and affective outcomes (Black & Wiliam, 1998; Boaler, 2002; Boaler & Greeno, 2000; Cotton, 2001; Lapointe, Legault, & Batiste, 2005; Meece, Herman, & McCombs, 2003; Ryan & Williams, 2007; Wilkins & Ma, 2003).

So under these three general theoretical perspectives and some identified gaps inherent in their interconnections (i.e. between teaching and learning outcomes), the TeLePriSM project aims to answer the following two general questions in regard to secondary mathematics education:

- **RQ1**: How can we measure (i) teachers' (self-reported) pedagogic practices and (ii) students' dispositions (and other learning outcomes) to study and use mathematics? And how do these measures vary across key subgroups (e.g. year groups), background variables (e.g. class, ethnicity, gender) and institutional types (school type)?
- **RQ2**: How do background and process variables (e.g. programme type) and pedagogy predict students' learning dispositions, outcomes and decisions from Y7 to Y11?

The particular research question we seek to answer in this paper regards measuring teaching practices. In particular, we set out to measure teaching practices from students' and teachers' point of view. Before describing our methods we summarise the conceptual framework around teaching practices next.

2. Conceptualising and Measuring 'teaching practices'

We detailed our conceptualisation of pedagogic practices rhetoric in our previous work (Pampaka, et al., 2012) – a brief summary of the relevant concepts is listed here:

- Literature on the intertwining of teaching and learning into 'a single entity' (Shuell, 1993; Vermunt & Verloop, 1999). Research on classroom learning environments has evolved to address this question, and found moderate positive associations between the learning environment and students' attitudes to mathematics.
- An extensive list of approaches presented as 'opposites' to 'transmission' model of teaching, usually as part of studies contrasting reform teaching with the existing dominant 'traditional' practice: 'facilitating' as against 'telling'; 'direct' versus 'indirect' instruction (Westerhof, 1992); 'scaffolding', 'individualised diagnostic learning approaches' (Bell, 1993); discussion-based approaches (Swan, 2000); guided discovery teaching approach; and 'dialogic mathematics teaching' (Ryan & Williams, 2007).
- Another common categorisation of classroom practices in the relevant literature is that between 'teacher-centred' and 'learner-centred' instruction or practices. Both are broadly applied to include a variety of views and strategies for teaching and learning (Cuban, 1983; Kember & Gow, 1994). The first, according to Schuh (Schuh, 2004), is usually associated with 'transmission' models of teaching where teacher and instruction are the focus, whereas 'learner-centred' practices move the focus to students and learning outcomes.
- There is, currently a lot of rhetoric that favours connectionist pedagogic practices at all levels of mathematics education in the UK (ACME, 2009; OFSTED, 2008). According to these documents, effective mathematics teaching should be connectionist in two senses: (a) connecting teaching to students' mathematical understandings, and productions (hence student-centred, but also involving assessment for learning, dialogic and discussion-based communicative mathematics); and (b) connecting teaching and learning across mathematics' topics, and between mathematics and other (e.g., scientific) knowledge.

The above theoretical perspectives will guide our findings' discussions as we embark on this project to link pedagogy with students' experiences and aspirations during their secondary education.

3. The TeLePriSM Project Design

The project design involves capturing five years of progression (Year 7 to 11) in one year of data collection, as shown with Figure 1, which poses methodological challenges around the combination of longitudinal and cross-sectional analyses[2].

[2] This refers to the third fundamental research question of the project, which goes beyond the scope of this paper

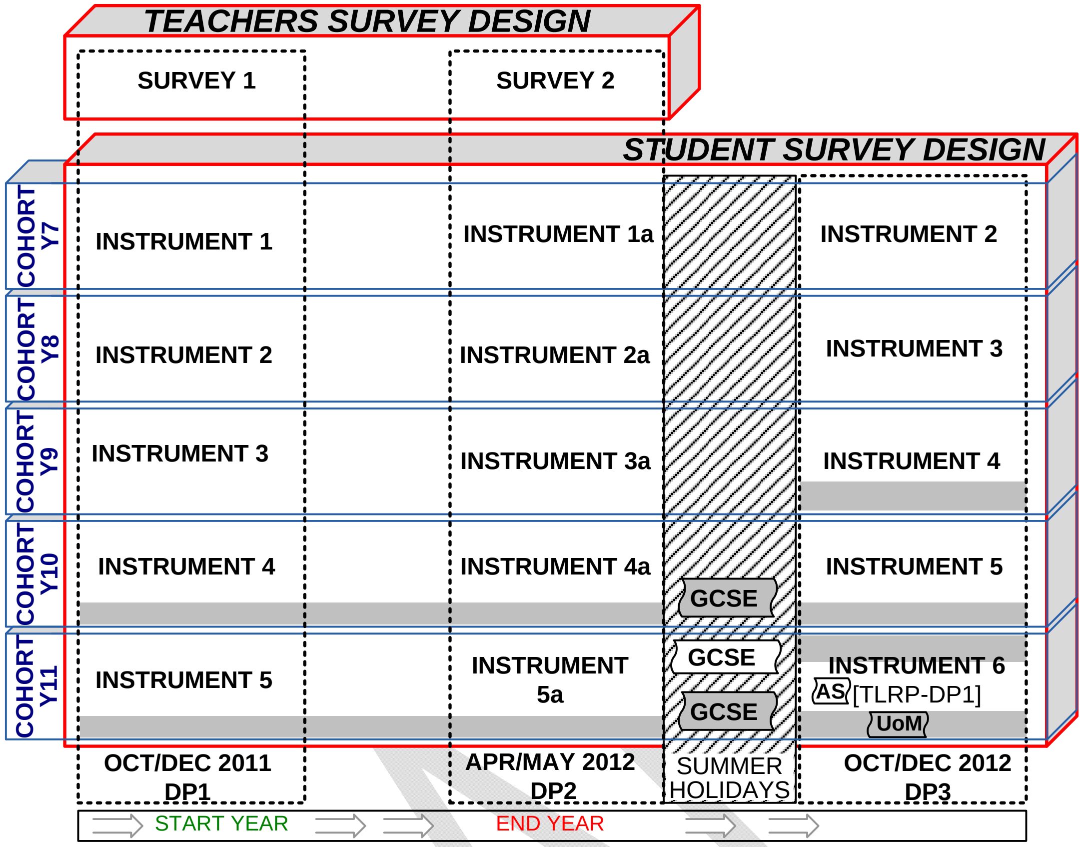


Figure 1: The TeLePriSM project design

The analysis and results presented here are only based on the first data point (DP1) which took place at the beginning of the previous academic year (2011-2012). Details about analytical approaches and the sampling are detailed in the following section.

4. Methodology – Analytical Framework

Our work is guided by an adaptation of a unified analytical/methodological framework for modelling social phenomena we already presented elsewhere (Pampaka, Williams, & Hutchenson, 2011). This version is expanded to include a forth step of 'resolving methodological challenges': Each step of the framework (Figure 2) addresses each of the three general research questions of the project: This will result in valid measures of teaching and learning as well as substantive results in regards to their variations across groups of teachers and students (RQ1), predictive models of students learning outcomes and dispositions to inform the mathematics community (RQ2), and practical applications of dealing with methodological challenges involved in dealing with complex designs (RQ3).

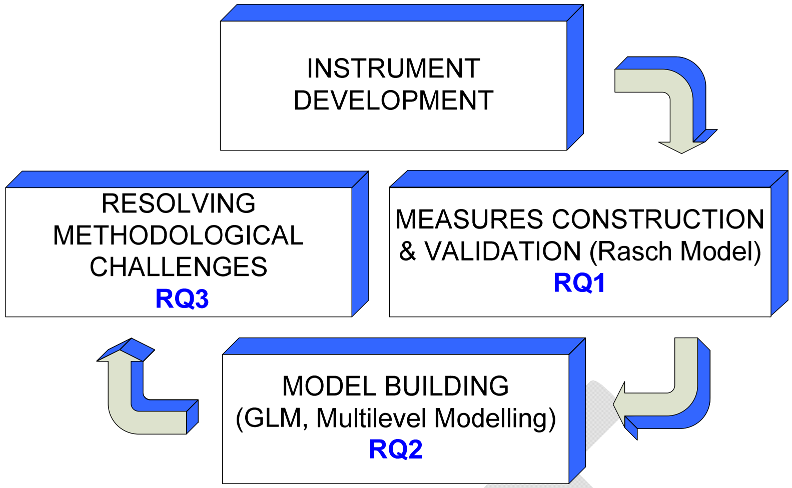


Figure 2: The TeLePriSM analytical framework

The methods and decisions taken around the three sections of the framework relevant to this paper are detailed next.

4.1. Instrument Development (and Piloting)

Teachers' Survey:

The teacher survey was largely influenced by our already validated version of a teacher instrument (Pampaka, et al., 2012) which we had previously developed building on the work of Swan (2006), which in turn built on the research findings of Askew et al. (Askew, et al., 1997) and Ernest (Ernest, 1991): Swan adapted three components that can be used to characterise the teachers' belief system (i.e., the nature of mathematics as a subject, the nature of mathematics teaching and the nature of the processes of learning mathematics). From the work of Askew and colleagues he derived the 'ideal' categories of teachers' orientation towards each component (i.e., transmission, discovery and connectionist). What we did and presented in detail in Pampaka et al (2012) was that we used the items of Swan's 'practice scale' (Swan, 2006) with some amendments and created a unidimensional measure of 'teacher centricism' or 'transmissionist' teaching of mathematics.

The 28 items of that instrument were revised (and some dropped) accordingly and complemented with items from other instruments reported in existing literature (Harwood, Hansen, & Lotter, 2006; Hiebert, et al., 2003; Kember & Gow, 1994; McCaffrey, et al., 2001; NCES, 2000; Roelofs, Visser, & Terwel, 2003; Swanson & Stevenson, 2002; Webster & Fisher, 2003) in order to reflect teaching practices found in 11-16 classrooms.

The teachers' (online) questionnaires were split into two parts: the first part was about the teacher asking for some background information and had to be completed only once. The second part was about their teaching with certain mathematics classes. The teachers were asked to complete a separate survey (Part 2) for each of the classes they teach and take part in the student survey.

The items for the second part of the survey are listed in Table 1. They were followed by the instruction "About how often do you do each of the following in your mathematics instruction in this class?", and each had a 4-option response: "Rarely", "Sometimes", "Often" and "Always".

Table 1: The items of the teacher survey (with reversed items highlighted)

teaching_1	I introduce a new topic by first determining what the students already know about it
teaching_2	I offer content matter in gradually increasing levels of complexity
teaching_3	I teach each topic from the beginning, assuming they know nothing
teaching_4	I teach the whole class at once
teaching_5	I jump between topics as the need arises
teaching_6	I have my students work collaboratively in pairs
teaching_7	I have my students work collaboratively in groups
teaching_8	I teach each student differently according to individual needs
teaching_9	I encourage students to discuss the mistakes they make
teaching_10	I tend to follow the textbook closely
teaching_11	Students work on projects in which subject material from various subjects is integrated
teaching_12	Students decide for themselves whether it is necessary to cooperate with other students
teaching_13	Students engage in mathematical activities using concrete materials
teaching_14	Students make formal presentations to the rest of the class
teaching_15	Students work on extended mathematics investigations or projects (a week or more in duration)
teaching_16	Students start with easy questions and work up to harder questions
teaching_17	Students read from a mathematics textbook in class
teaching_18	Students use mathematical concepts to interpret and solve applied problems
teaching_19	Students play mathematics games
teaching_20	Students work through exercises from textbooks or worksheets
teaching_21	Students work on their own, consulting a neighbour from time to time
teaching_22	Students choose which questions to tackle
teaching_23	I choose examples that appeal to students
teaching_24	I try to indicate the value of each lesson topic for future use
teaching_25	When a student asks a question, I give a clue (or scaffold) instead of the correct answer
teaching_26	During instruction I ask a lot of short questions to check whether students understand the content matter
teaching_27	I assign mathematics homework
teaching_28	I ask students to explain their reasoning when giving an answer
teaching_29	I encourage students to explore alternative methods for solutions
teaching_30	I allow students to work at their own pace

Students' Survey:

Student surveys are based on different versions of the same instrument (as shown in Figure 1) to reflect the age and level of students. Background variables and measures of students' attainment are also being collected including gender, ethnicity, language of first choice, proxies of socioeconomic status, and earlier NC level records and GCSE grades. The various sections of the questionnaire capturing teaching and learning perceptions have been constructed and expanded based on our TransMaths framework:

■ *Dispositions and Self Efficacy*: Instruments for mathematics self efficacy and dispositions towards further maths study have previously been developed, calibrated and validated for 16 year old students (Pampaka, Kleanthous, Hutcheson, & Wake, 2011; Pampaka, Williams, Davis, & Wake, 2008; Pampaka, et al., under review; Pampaka, et al., 2012). These instruments have been then modified to suit the age range Y7-11 including new items as appropriate (Gwilliam & Betz, 2001; Nielsen & Moore, 2003; Pietsch, Walker, & Chapman, 2003). It should be noted that the main difference across the different versions of the surveys regarded the task-specific self-efficacy items.

■ *Students' perception of pedagogy:* This instrument which is the focus of this paper, had been developed, linking as well to our previous work with teacher instruments and adopting existing measures of perceived classroom environment (Aldridge, Fraser, Taylor, & Chen, 2000; Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008; Fraser, 1998).

The instrument we used for the main study included the 26 items presented in Table 2. These were forming the main part of Part D of the Questionnaire (entitled How Maths is taught and learnt) under the instruction: "Please tell us, how often does the following happen in your maths lessons?"

After this list of statements students were also asked to rate the overall perceived difficulty of their maths lessons (too easy, about right, too hard), and report in an open question what they use computers and calculators for.

Refining and Piloting the Instruments

Different refinement stages took place between March 2011 until the actual online piloting in June/July 2011. These included a lot of checks among the team and discussions with expert researchers, as well as teachers. Various versions were also piloted in the form of 'interviewing' small numbers of students in some schools. A refined version for each was then developed into an online version, which was used for pilot data collection at two schools during summer of 2011. The purpose of the pilot was two-fold: (a) to check the stability of the online tool and other technical issues arising from surveying online students in their classes, (b) to collect data to validate the different measures we were hope to develop.

Table 2: The items of the instrument for students' perception of pedagogy

[Please circle the appropriate number in each line]		Never	Rarely	Sometimes	Always
1	The teacher asks us questions.	1	2	3	4
2	The teacher asks us to explain how we get our answers.	1	2	3	4
3	The teacher starts new topics with problems about the world.	1	2	3	4
4	The teacher tells us to work more quickly.	1	2	3	4
5	The teacher uses the computer to teach some topics.	1	2	3	4
6	The teacher gives us problems to investigate.	1	2	3	4
7	The teacher expects us to remember important ideas we learned in the past.	1	2	3	4
8	The teacher tells us which questions/activities to do.	1	2	3	4
9	The teacher asks us what we already know about a lesson topic.	1	2	3	4
10	The teacher tells us what value the lesson topic has for future use.	1	2	3	4
11	We work together in groups on projects.	1	2	3	4
12	We listen to the teacher talk about the topic.	1	2	3	4
13	We copy the teacher's notes from the board.	1	2	3	4
14	We talk with other students about how to solve problems.	1	2	3	4
15	We ask other students to explain their ideas.	1	2	3	4
16	We do projects (assignments) that include other school subjects.	1	2	3	4
17	We work through exercises from the textbook.	1	2	3	4
18	We learn how mathematics has changed over time.	1	2	3	4
19	What we learn is related to our out-of-school life.	1	2	3	4
20	We learn that mathematics is about inventing rules.	1	2	3	4
21	We get assignments to research topics on our own.	1	2	3	4
22	We use calculators.	1	2	3	4
23	We use computers.	1	2	3	4
24	We use other things like newspapers, magazines, or video.	1	2	3	4
25	We discuss ideas with the whole classroom.	1	2	3	4
26	We explain our work to the whole class.	1	2	3	4

4.2. Measure Validation

The validation process refers to the accumulation of evidence to support validity arguments regarding both teachers' pedagogic measure as well as students' disposition measures. Our psychometric analysis for this purpose is conducted within the Rasch measurement framework and therefore we follow the relevant proposed guidelines (Wolfe & Smith Jr., 2007a, 2007b) based on Messick's definitions of validity (Messick, 1988, 1989; Silva, 1993). The Rasch model has been selected because it provides the means for constructing interval measures from raw data and because the total raw score is sufficient for estimation of measures (Wright, 1977). Models of the Rasch family are governed by certain assumptions, the most important of which are unidimensionality, local independence, and common item discrimination. In its simplest form (i.e. for dichotomous responses) the model proposes a mathematical relationship between a person's 'ability', the 'difficulty' of the task, and the probability of the person succeeding on that task. The Rasch model in this case help to construct simple, fit for purpose, one-dimensional measures. We have been extensively employing this approach for the validation of our constructed measures, and had already reported some of these findings elsewhere (Pampaka, Kleanthous, et al., 2011; Pampaka, Williams, et al., 2011; Pampaka & Williams, 2010; Pampaka, et al., under review; Pampaka, et al., 2012). The Rasch rating scale model is considered the most appropriate for the scaling problems we have in this particular paper (i.e. a common frequency scale) (Bond & Fox, 2001). Our decisions about the validity of each measure are based on the following statistical indices (which we derive from data analysis in Winsteps):

(i) *Item fit statistics* to indicate how accurately the data fit the model, and thus provide evidence for fulfilment of the unidimensionality assumption, hence suggesting development of one-dimensional scales. Misfit could then suggest the possibility of existence of new dimensions which we will examine further, by employing multidimensional models (Raudenbush, Johnson, & Sampson, 2003).
(ii) *Category Statistics* will be examined for the appropriateness of the Likert scale used and its interpretation by the respondents, to justify what is usually called as communication validity (Linacre, 2002; Lopez, 1995, 1996).
(iii) *Person – item maps and the item difficulty hierarchy* will provide evidence for substantive, content and external validity.
(iv) *Differential Item Functioning (DIF) and person fit statistics* will suggest group differentiation of the constructed measures, which is an important aspect of validity when an instrument is used with different groups of persons (Thissen, Steinberg, & Wainer, 1993). DIF will be examined in this study for gender (student sample) and year group (for both samples).

4.3. Further Statistical Modelling

The design of Figure 1, leads to datasets from various sources. A straightforward distinction resulting directly from the surveys leads to student level, and teacher level data. However, as will be explained later on, we also have data at classroom level. On another perspective we have longitudinal matched data or cross-sectional data for each DP: since now we focus on the first DP we avoid this complication for the time being. So, analysis can be performed in various ways. For the purposes of this paper we will focus on ‘simple’ analysis.

- Teachers’ surveys will be analysed initially as a separate dataset to compare pedagogic practice across classes in the 5 year groups
- Students’ responses to the previously detailed ‘pedagogical’ instrument will also be analysed separately.
- The students’ dataset will then be complimented with the corresponding maths teachers’ pedagogic measures in order to first check the ‘agreement’ between the scores and investigate some preliminary patterns and associations regarding pedagogy.
- Linking backwards (to class level data) will also be employed for this analysis to further explore patterns of agreement between teachers’ scores and average measures of their students’ scores.

In order to clarify more the above datasets, the sampling procedures and resulting sample are presented next.

5. Sampling & Sample(s)

5.1. The TeLePriSM Sampling Design

The nature and design of the study (i.e. longitudinal at school level for selection purposes) make it necessary to employ a varied sampling frame to ensure maximum coverage. The original plan was to enlist at least 50 schools representing the range of schools in England. We accepted early on that aiming for a representative sample of schools across the country is not a feasible task, since studies of this kind are largely limited to self-selected samples[3]. Therefore we invited schools, drawing on the following sources:

- School contacts already known by us or other colleagues (e.g. PGCE courses, qualification agencies)
- A schools’ database we purchased from a private company: we decided to approach schools within 30 miles of various cities (to get a variety of urban/rural schools) across the country. The cities chosen were: Manchester, London,

[3] However we have plans in place to investigate the comparability of our sample to the national one

Birmingham, Leeds, Bristol, Newcastle, Cambridge and Oxford (only 10 miles to avoid overlap with earlier selections).

The initial requirement was for schools to take part with all their Year 7 to 11 mathematics teachers and classes and be willing to follow this up at 2 more data collection points. In total, we approached over 2200 schools and we were able to establish collaboration with 40 of them, spread geographically as shown below:

Figure 3: The distribution of participating schools

5.2. Our resulting samples (at DP1)

As already mentioned, this analysis draws on the first data point (out of three) of our longitudinal survey of students in Year 7 to 11 and their mathematics teachers. This data collection took place from October to December 2011.During this period we collected data from a sample of more than 13,000 students in 40 schools. Some further information about school description according to gender composition, and age range is given in Table 3, whereas Table 4 shows the total number of students per year group in the ***students' dataset***.

Age range	Boys only	Girls only	Mixed	Total
11-16	0	2	13	**15**
11-18	1	5	19	**25**
Total	1	7	32	**40**

Table 3: Basic types of schools

Year 7	3884
Year 8	3025
Year 9	2668
Year 10	2145
Year 11	1794
Total	**13516**

Table 4: Total number of students per year group in dataset

Teachers' dataset: 128 individual teachers completed the teacher questionnaires for some of their classes. This resulted in 264 surveys of classroom practice corresponding to 5062 students of our sample.

Class-Teacher Matched dataset: From students' responses it was possible to identify a total of 762 classes which we use as a unit of analysis for some results in this paper (this corresponds to 13491 students of the total 13516). This dataset includes the 264 matched cases for which we had a completed teacher survey.

6. Validation Results

In this section we deal with the procedures employed for ensuring construct validity and the presentation and interpretation of the two constructed measures of 'pedagogy' from the two separate analyses.

6.1. Teacher's Self Reported Pedagogy

These results regard the Rasch Rating Scale model analysis of the 264 cases of pedagogy as reported by the 132 teachers.

Some preliminary analysis of all items without recoding was performed to investigate how they map: these results may be suggestive of multidimensionality an issue that is beyond the scope of this paper and will be explored further at a later stage.

For this first analysis we follow the procedures established in our previous work, which hugely informs the methods and approaches for this project (i.e. Pampaka, et al., 2012). In order to establish a unidimensional model with some meaning and maintain the direction of the measures' intensity the reversal of the original coding of some items was necessary: these items are highlighted in Table 1 and their reversed coding should be taken into account in the interpretation of the constructed scale. Some statistical indices are analysed below as suggested by the measurement framework we follow.

Item fit statistics

Initially we check the item fit statistics, which provide an indication of how accurately the data fit the model, in the Rasch context. Inconsistent data may suggest the existence of new dimensions in the data, hence lack of fulfilment of the unidimensionality assumption in a worst case scenario, or could simply become a source of further inquiry. Fit statistics may also flag items to which responses are overly predictable (overfits), an indication that, in some way, they are dependent on the other items and might be the first choices for deletion (Bowles, 2003; Wright, 1994). There is a debate about setting cut-off points for acceptable ranges of fit statistics (Linacre, 2002; Smith, Schumacker, & Busch, 1998), which we acknowledge and thus consider existing recommendations regarding the format of our data and take the value of 1.3 as a value for infit and outfit mean squares that suggests cause of concern. Particularly we will refer to items with fit statistics higher than this value as 'misfits' and use this as an indicator of items worth further (possibly more qualitative) investigation. Having said this, we also endorse Bohlig et al.'s (1998) recommendation that 'less than pleasing fit statistics say "think again", not "throw it out"' (p. 607), and hence we seek explanations and interpretations for the high fit values, and we only disregard the misfit items from our scales if there is a 'good' reason.

Preliminary analysis of all 30 items indicated some problems with the first item so it was removed from the measurement scale (the main reason for this beyond fit statistics is the similarity of the content of the item with Item 3 in this scale).The Rasch analysis of the remaining 29 items showed acceptable fit overall of almost all the items, and this supports the assumption of the existence of a unidimensional scale. In other words, we can claim that they measure what we call teacher self-report perception of a transmissionist pedagogic practice'. The results (fit statistics and item measures) are presented in Table 5.

Some slightly misfitting items are highlighted in Table 5 all presenting infit values of higher than 1.3. Item 3: "I teach each topic from the beginning, assuming the know nothing", and Item 17: "Students read from a mathematics textbook in class" are both items from the 'difficult' end of the scale (as given by the high measure scores they have). In other words these are items that are the hardest to report frequency of occurrence by any teacher and maybe their slight misfit is due to some outlier values. Similarly the reversely coded Item 15 is also misfitting and also the easier item on the scale: its interpretation should consider the reversal, thus the negation: "Students (don't) work on extended mathematics investigations or projects". Maybe these items are indicating differences among different groups (this will be partly explored with DIF) but because of their positioning and their role in defining the two ends of the scale (as will also be shown with the person-item map later) these items are sustained in this measure.

Table 5: Item statistics for measure of reported 'pedagogical practice'

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
1	899	262	-1.47	.10	.88	-1.6	.88	-1.6	.25	.22	53.4	51.6	teaching_2
2	499	264	1.69	.09	1.36	4.4	1.35	4.2	.39	.25	43.6	48.8	teaching_3
3	802	262	-.57	.09	.83	-2.2	.82	-2.2	.40	.24	58.0	55.2	teaching_4
4	717	262	.10	.09	1.13	1.5	1.12	1.4	.20	.26	50.8	52.3	teaching_5
5	678	263	.40	.08	1.04	.5	1.06	.7	-.28	.26	41.1	49.0	teaching_6
6	786	263	-.41	.09	1.08	.9	1.08	.9	.47	.25	48.3	55.4	teaching_7
7	690	263	.31	.09	1.08	1.0	1.08	1.0	.25	.26	48.3	50.0	teaching_8
8	492	264	1.74	.09	1.05	.7	1.04	.5	.27	.25	49.2	48.5	teaching_9
9	409	263	2.43	.10	1.25	2.8	1.26	2.9	.25	.22	49.0	50.2	teaching_10
10	890	261	-1.41	.10	1.20	2.3	1.16	1.9	.37	.22	50.6	51.8	teaching_11
11	765	262	-.27	.09	1.30	3.3	1.31	3.4	.22	.25	43.9	55.0	teaching_12
12	740	261	-.10	.09	.72	-3.7	.71	-3.7	.38	.25	66.7	54.1	teaching_13
13	911	263	-1.56	.10	1.20	2.3	1.18	2.1	.33	.21	55.5	53.1	teaching_14
14	923	261	-1.78	.11	1.32	3.5	1.22	2.4	.43	.21	60.5	57.0	teaching_15
15	828	262	-.79	.09	.75	-3.2	.75	-3.4	.38	.24	63.7	54.1	teaching_16
16	384	261	2.65	.10	1.34	3.5	1.35	3.5	.11	.21	53.3	56.8	teaching_17
17	624	263	.78	.08	.72	-4.0	.72	-4.0	.21	.26	54.0	47.1	teaching_18
18	714	261	.10	.09	.85	-1.8	.86	-1.7	.38	.26	52.1	52.3	teaching_19
19	698	262	.24	.09	.78	-2.9	.77	-3.0	.42	.26	60.3	50.9	teaching_20
20	693	262	.27	.09	.69	-4.2	.68	-4.3	.33	.26	63.4	50.3	teaching_21
21	770	262	-.31	.09	1.19	2.2	1.19	2.2	.27	.25	49.2	55.2	teaching_22
22	744	264	-.06	.09	1.06	.8	1.08	.9	-.16	.25	52.7	53.7	teaching_23
23	556	263	1.26	.08	.85	-2.0	.85	-2.1	.28	.26	55.5	48.7	teaching_24
24	892	264	-1.33	.10	1.02	.3	1.03	.5	.12	.22	50.0	51.1	teaching_25
25	862	263	-1.07	.10	.93	-.9	.94	-.7	.18	.23	53.2	52.0	teaching_26
26	907	264	-1.48	.10	1.13	1.5	1.14	1.7	.07	.22	48.5	51.5	teaching_27
27	895	264	-1.36	.10	1.08	1.0	1.11	1.4	-.11	.22	45.1	51.4	teaching_28
28	588	263	1.03	.08	.74	-3.6	.74	-3.6	.33	.26	55.9	47.7	teaching_29
29	598	264	.98	.08	.81	-2.6	.81	-2.6	.22	.26	54.5	47.5	teaching_30
MEAN	722.6	262.6	.00	.09	1.01	.0	1.01	.0			52.8	51.8	
S.D.	151.4	1.0	1.19	.01	.21	2.6	.20	2.5			6.1	2.7	

Response category statistics

Categorization is crucial in designing any ordered-response scale (including the rating scale) and it has two important characteristics: (a) while all categories of a scale should measure a common trait or property, each of them must also have its own well-defined boundaries, and the elements in a category should all share certain specific exclusive properties, and (b) categories must be in an order, and numerical values generated from the categories must reflect the degrees or magnitudes of the trait.

Category statistics are given as indicators for this check: the most frequently used indices are the average measure and the threshold (or step calibration). A well functioning scale should present ordered average measures, with acceptable fit statistics, as shown for our case in the results of Table 6. It should also present ordered step calibrations, which is also supported with Table 6 and the category probability curves of Figure 4.

Table 6: Summary of category structure

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	770	10	-1.25	-1.23	1.00	1.03	NONE	(-2.92)	1
2	2	1991	26	-.18	-.20	.99	.99	-1.67	-.98	2
3	3	3218	42	.76	.78	.99	.99	-.18	.91	3
4	4	1637	21	1.55	1.53	1.01	1.03	1.85	(3.04)	4
MISSING		40	1	.75						

```
        CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P      -+-------+-------+-------+-------+-------+-------+-------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                       44|
A      |                                                    444  |
B   .8 +1                                                 44     +
I      | 11                                            444       |
L      |   1                                         4           |
I      |    11                                     44            |
T   .6 +      11                                 44              +
Y      |        1                  3333333333    4               |
    .5 +         1   22222222    33          33 4                +
O      |          **2        2233             4*3                |
F   .4 +        22  1        3322            4   33              +
       |      22     11     3    22        44      33            |
R      |    22         1  33        22     4          33          |
E      | 222            1*          2  44              33        |
S   .2 +2              33 1           **2                 333     +
P      |            333    11        44   22                 333  |
O      |         333         111 444        222                 33|
N      |    33333          44444*11111        222222             |
S   .0 +****444444444444444           11111111111111*************+
E      -+-------+-------+-------+-------+-------+-------+-------+-
       -3      -2      -1       0       1       2       3       4
        PERSON [MINUS] ITEM MEASURE
```

Figure 4: Category probability curves

Differential Item Functioning (DIF) by Year Group

DIF could indicate group differentiation of the constructed measures, and because of this, it is an important aspect of validity when an instrument is used with different groups of persons (Thissen, et al., 1993; Wright & Masters, 1982). In this analysis, it is useful to check whether the items ‘function’ similarly across the 5 year groups of reported teaching practices. Some issues may be highlighted with the following two plots: Figure 5 reports the difficulty of the items for each group, whereas Figure 6 reports the size of the item DIF for each group relative to the overall item difficulty.

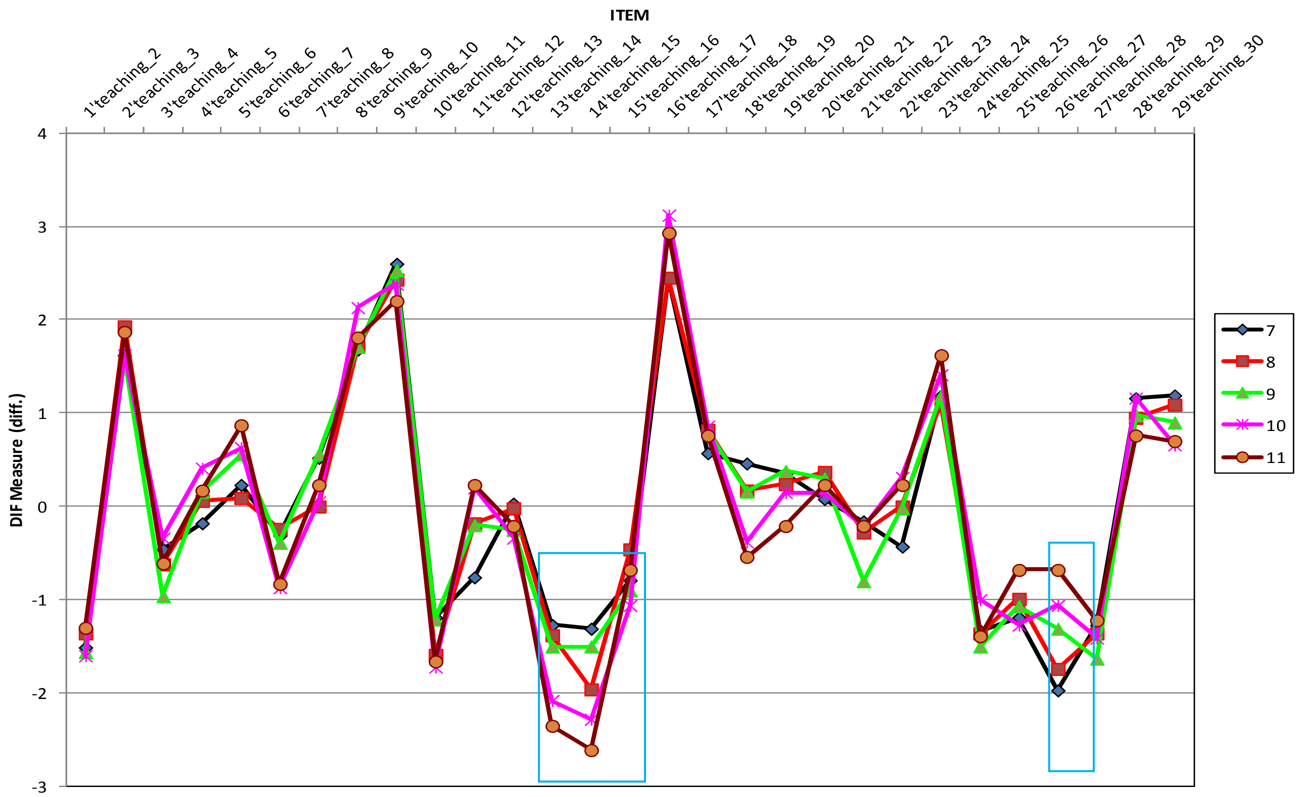


Figure 5: Item difficulty measures for each year group

As shown in Figure 5 the bigger variations[4] in the item difficulties appear to be for items 14 and 15, as well as item 27. The size of the DIF is better shown in Figure 6, which highlights the following patterns:

- As we move from year 7 to year 11 the frequency of teachers' assigning homework becomes smaller (given the higher difficulty score of item 27 for older year groups). This trend is linear (Year 7>Year8>Year9>Year10>Year11).
- The opposite trend seems to appear for items 14 and 15 (Students (don't) make formal presentations to the rest of the class, and students (don't) work on extended mathematics investigations or projects). Based on this trend and reversed coding we could infer that as we move on to older year groups, formal presentations and extended investigations become a less frequent practice.

[4] We also consider some guidelines that suggest that DIF size more than 0.5 logits is a cause of concern

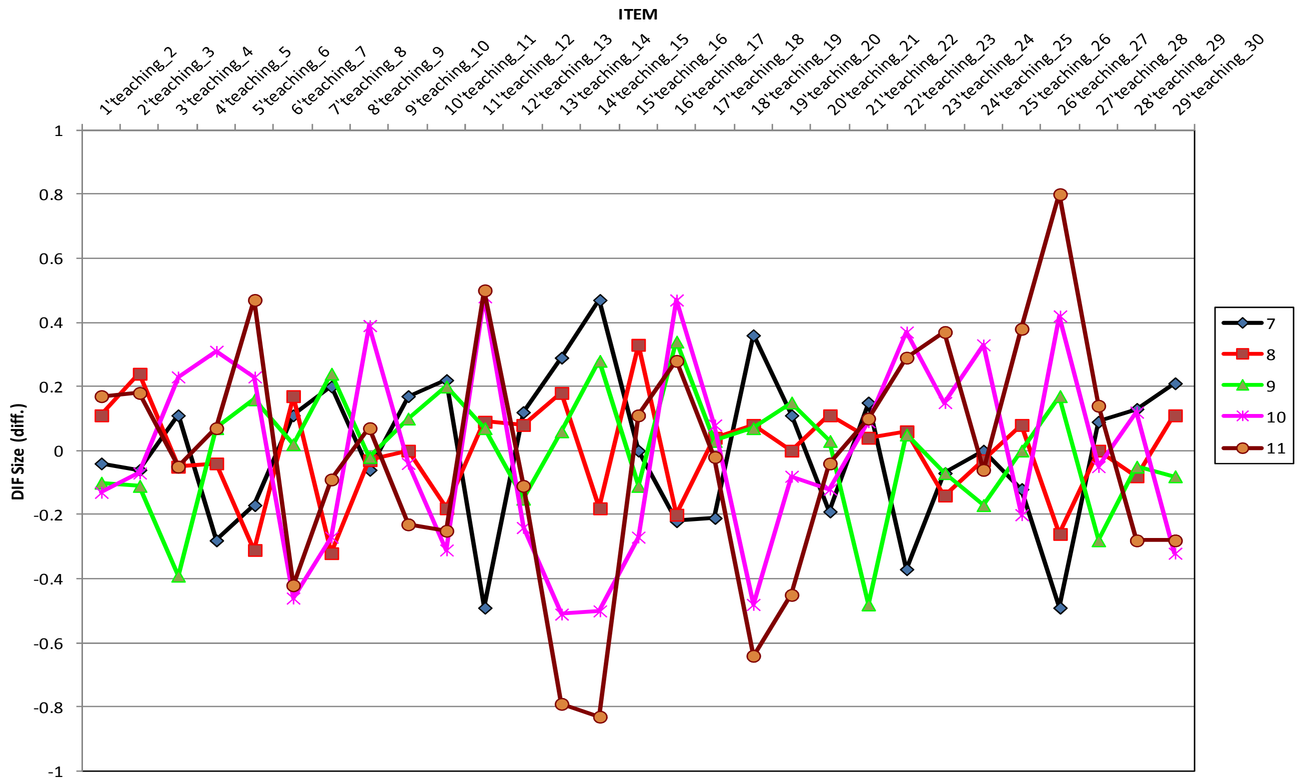


Figure 6: DIF size (compared to overall) for each year group

The resulting measurement scale

Figure 7 on the left shows the resulting measurement scale, with the items listed on the right (with the reversed ones highlighted). The measurement scale is presented by a person-item map which is one of the unique features of the Rasch model: it plots both items and persons on a common scale. The unit for this scale is the logit (which can be transformed in a more meaningful range, but for this purpose we analyse the original result). On the left of the item-person map, the distribution of the pedagogic measures for each teacher-class case is shown (as a histogram). The higher the place on the histogram, the more teacher-centred or transmissionist the pedagogy. Pedagogy that is mainly student-centred or connectionist is (or should have been) at the bottom. On the right hand side of the map the items that constitute the scale are presented, ranging from those easiest to report as frequent to the most 'difficult' to report being frequent. For reversed items the opposite happens, so they are a negated in the figure for easier interpretation.

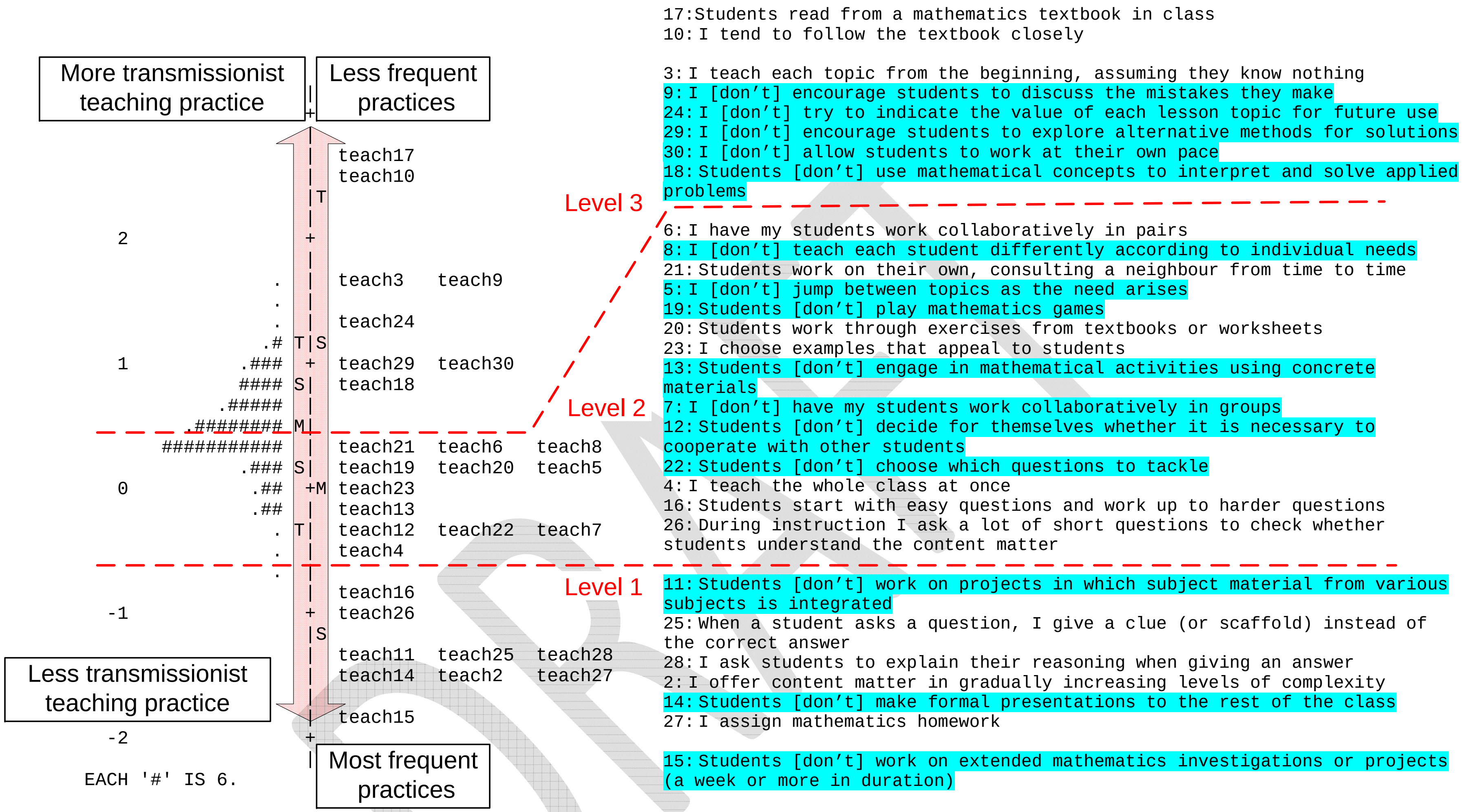


Figure 7: The person-item map and the resulting 'transmissionist' pedagogy scale

Three hierarchical categories of this ‘teacher centredness’ measure can perhaps be distinguished (as shown in Figure 7, by the dotted lines) based on both the statistical results (i.e., item and person separation, Table 3) and the qualitative analysis of the homogeneity of the item content, similarly to our previous work (Pampaka, et al., 2012). Based on this preliminary categorization, the teachers-classes practices can also be split into the three levels defined in Figure 7: Level 1 teachers’ practice is frequently student-centred and more connectionist; Level 2 involves teachers’ practices from both ends of the spectrum in moderate frequencies and Level 3 corresponds to more transmissionist, teacher-centred practices. It should be emphasised that these classifications are just preliminary and further work is needed for their justification also involving qualitative evidence from teacher and student interviews. One observation, though, that seems to be consistent and could be considered as the first finding from this analysis is that we did not find a lot (or at all) non-transmissionist teacher-classes cases so far. The majority of the reported practices, as also shown with the histogram, are skewed towards the top end of the scale.

6.2. Students’ Perception of the Pedagogy they Experience

A similar procedure was followed for the student sample (N=13516): the 26 items listed in Table 2 were analysed employing the Rasch Rating Scale model, considering also the reversal of some items so as the resulting measure will be of the same direction as the teacher scores, and our previous findings. The main results will be presented briefly in this section since most of the statistical indices and their interpretation were explained in detail during the presentation of the teacher reported measures.

As shown in Table 7 the items work very well together to define this measure of we call ‘students’ perception of transmissionist pedagogy’. None of the item exhibit fit above the acceptable range so there are no major causes for concern from this initial analysis. The fit statistics also indicate a well-behaved rating scale.

Figure 8 shows the DIF size for the items for each Year Group. The differences are small in regards to their logit size (but considerably big if we consider the range of this measurement scale). The biggest differences are briefly discussed and further analysis will be considered to resolve any potential problems. As shown, items 21 and 22 first, and then item 16 exhibit differences all at the same direction. Doing projects/assignments on their own (item 21) or involving other school subjects (item 16) occur in much less frequency as students move from Year 7 to Year 11 (those items were reversely coded). On the contrary, and probably as expected, the use of calculators is more frequent in Years 10 and 11.

Table 7: Item statistics for measure of 'perceived pedagogical practice'

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
1	47644	13060	-1.49	.02	1.07	4.5	1.08	4.6	.23	.21	69.6	67.5	lessons1
2	45741	13011	-1.11	.01	1.00	-.1	1.05	3.3	.17	.24	53.2	54.4	lessons2
3	36117	12866	.19	.01	1.11	9.9	1.12	9.9	.40	.31	37.1	43.0	lessons3
4	36148	12970	.22	.01	.97	-2.9	.97	-2.4	.17	.31	47.5	42.9	lessons4
5	39481	12973	-.18	.01	1.08	6.7	1.13	9.9	.05	.29	52.3	43.8	lessons5
6	39929	12949	-.24	.01	.96	-4.0	1.01	1.0	.02	.29	56.8	43.9	lessons6
7	42968	12966	-.66	.01	.96	-3.0	.99	-.8	.21	.27	49.9	46.0	lessons7
8	47479	12937	-1.57	.02	1.14	8.2	1.09	5.5	.30	.20	73.7	69.5	lessons8
9	40497	12878	-.35	.01	1.14	9.9	1.22	9.9	-.01	.28	49.4	43.9	lessons9
10	30956	12828	.77	.01	1.08	7.7	1.09	8.3	.40	.31	44.0	43.1	lessons10
11	32504	12888	.61	.01	.74	-9.9	.75	-9.9	.44	.31	51.1	42.6	lessons11
12	44006	12855	-.89	.01	1.04	2.9	1.09	6.8	.15	.25	49.5	49.2	lessons12
13	40896	12847	-.41	.01	.93	-6.3	.95	-4.3	.21	.28	51.8	44.1	lessons13
14	25718	12824	1.39	.01	.87	-9.9	.87	-9.9	.27	.30	57.2	45.0	lessons14
15	27191	12807	1.20	.01	.91	-8.4	.91	-8.3	.29	.31	53.4	44.8	lessons15
16	36534	12778	.11	.01	.92	-7.3	.92	-7.3	.57	.30	39.4	43.2	lessons16
17	38491	12800	-.12	.01	1.19	9.9	1.22	9.9	.18	.30	47.6	43.8	lessons17
18	37365	12787	.02	.01	1.00	.3	1.00	-.3	.57	.30	36.1	43.5	lessons18
19	32611	12734	.56	.01	.94	-6.2	.94	-5.9	.52	.31	45.2	42.6	lessons19
20	33788	12684	.41	.01	.98	-2.1	.98	-1.6	.50	.31	41.3	42.5	lessons20
21	35695	12672	.18	.01	.97	-2.5	.98	-2.3	.57	.31	37.0	43.0	lessons21
22	36120	12839	.18	.01	.79	-9.9	.81	-9.9	.11	.31	59.4	43.0	lessons22
23	35816	12817	.21	.01	.95	-4.9	.95	-4.7	.39	.31	42.8	42.9	lessons23
24	42787	12784	-.72	.01	.96	-3.0	.93	-5.3	.44	.26	50.7	46.8	lessons24
25	24275	12806	1.57	.01	1.07	5.6	1.08	6.6	.20	.29	49.2	44.7	lessons25
26	36456	12718	.11	.01	1.21	9.9	1.26	9.9	-.07	.30	48.5	43.2	lessons26
MEAN	37200.5	12849	.00	.01	1.00	-.2	1.01	.5			49.8	46.3	
S.D.	6009.0	97.2	.77	.00	.11	6.8	.12	6.9			8.8	6.9	

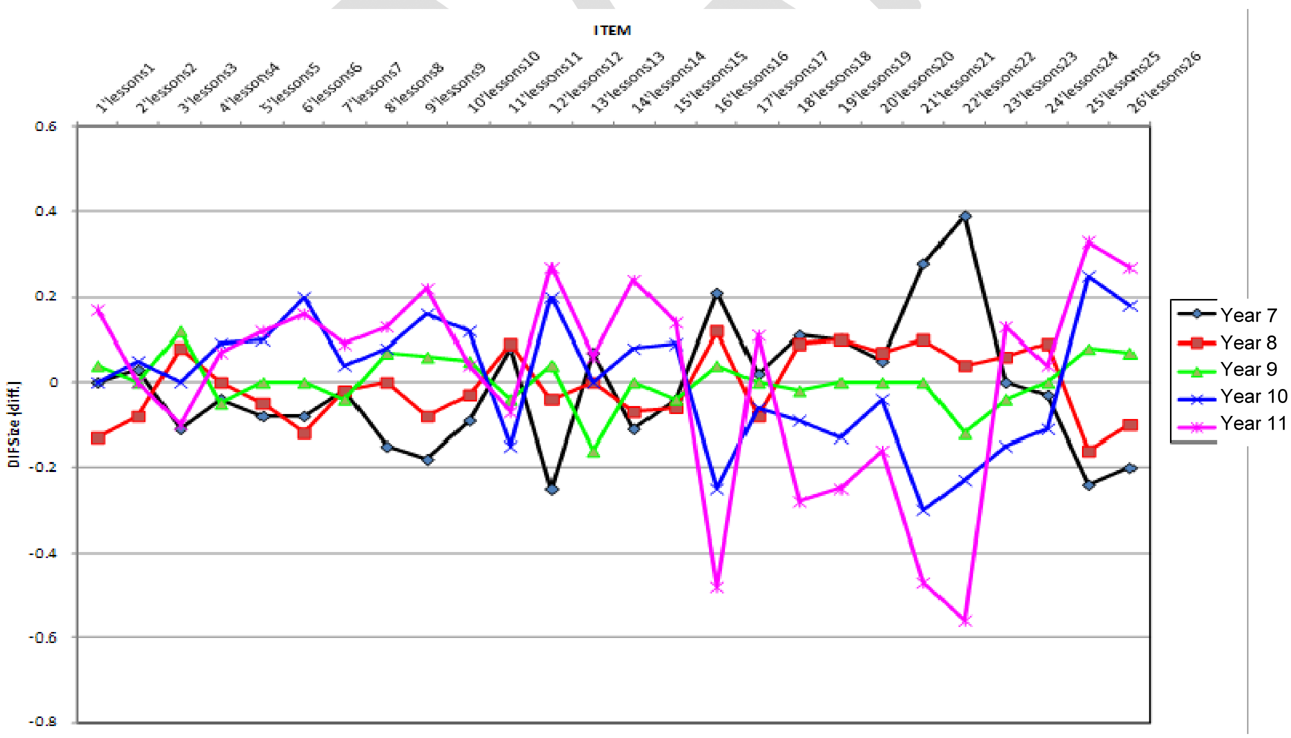
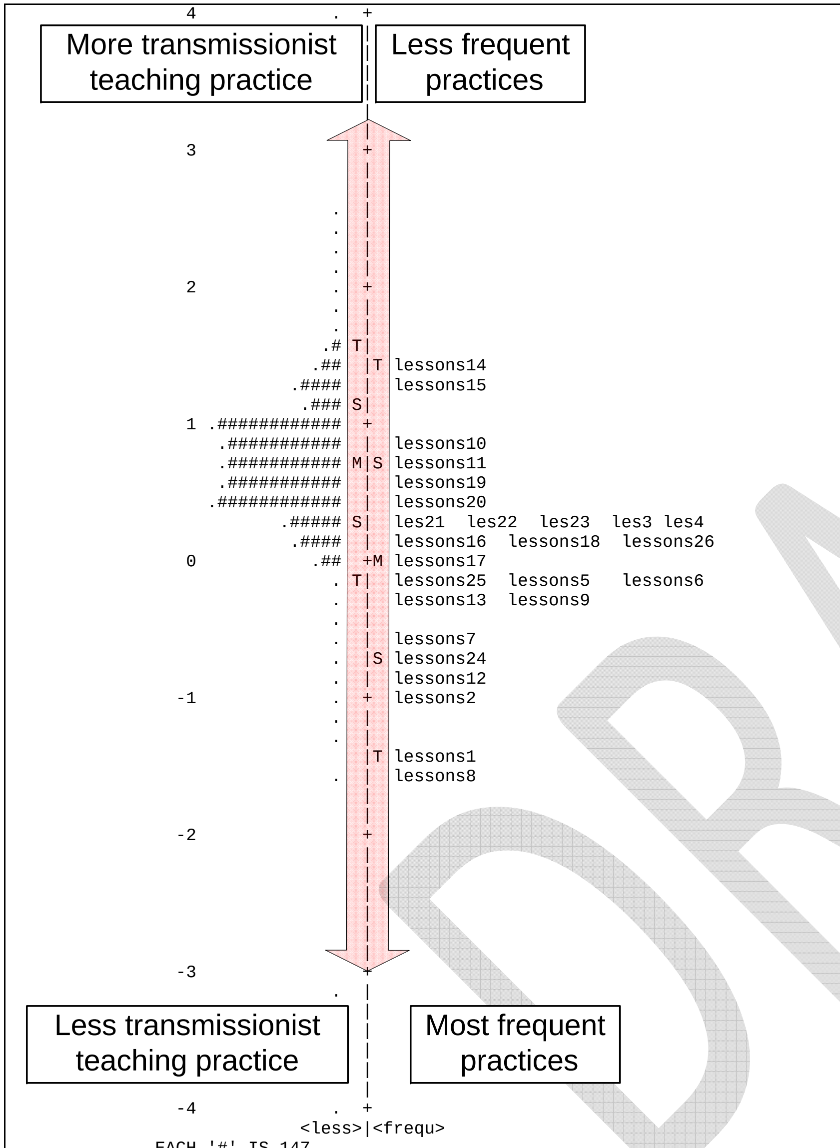


Figure 8: DIF size (compared to overall) for each year group (students)



14: We talk with other students about how to solve problems.
15: We ask other students to explain their ideas.

10: The teacher tells us what value the lesson topic has for future use.
11: We work together in groups on projects.
19: What we learn is related with our out-of-school life.
20: We learn that mathematics is about inventing rules.
21: We get assignments to research topics on our own.
22: We use calculators.
23: We use computers
3: The teacher starts new topics with problems about the world.
4: The teacher tells us to work more quickly.

16: We do projects (assignments) that include other school subjects.
18: We learn how mathematics has changed over time.
26: We explain our work to the whole class.
17: We work through exercises from the textbook.
5: The teacher uses the computer to teach some topics.
6: The teacher gives us problems to investigate.
25: We discuss ideas with the whole classroom.
9: The teacher asks us what we already know about a lesson topic.
13: We copy the teacher's notes from the board.

7: The teacher expects us to remember important ideas learnt in the past.
24: We use other things like newspapers, magazines, or video.
12: We listen to the teacher talk about the topic.
2: The teacher asks us to explain how we get our answers.

1: The teacher asks us questions.
8: The teacher tells us which questions/activities to do.

Figure 9: The person-item map and the resulting students' perception of transmissionist pedagogy scale

Finally, Figure 9 shows the resulting measurement scale, with the items listed on the right (with the reversed ones highlighted). Its interpretation is similar to that of Figure 7, so it will not be repeated here. A similar concluding observation is worth stating though: As with the teachers' reports of pedagogy, the students' perceptions of what is happening in these classes gives a similar skewed distribution towards the transmissionist ends of the scale. These issues will be further explored in the next section.

6.3. Investigating the 'agreement' between teachers and students scores

The question of how much students' perception of pedagogy agree with those of their teachers is important for this study, and presents an extra validation check for this analysis. In order to come to a preliminary conclusion to this question two examinations were performed: one employing the student level data and the other with the class-teacher matched dataset briefly described earlier. It should be mentioned that both methods have limitations so their complimentary use will increase the evidence for claims made at least to a point.

At student level analysis each student was allocated the score of their teacher, if the teacher had completed a survey for that class. This was in addition to the other student-level variables already in the dataset, including students' scores on their 'perception of transmissionist' teaching. The agreement of the scores in this case was checked with Pearson correlations and was find to be positive but weak (r=0.095). This correlation was statistically significant (p-value = 2.578e-11, df=4959), which supports partly the agreement of the scores: as teachers' scores increases so does their students. We should emphasise here the hierarchical structure of the data and the need for multilevel modelling of these associations, which is our intention to explore in the future.

Another analysis involved the ***Class-Teacher Matched dataset*** which included 762 classes, from which 264 included the teacher reported scores. In order to get a measure of students' perception to this dataset another compromise had to take place: we had to use the average of students' scores for each class. To get more insight into these associations we also recorded for each class, the minimum, maximum and Standard deviation of the students' scores. We also correlated students' average ability in mathematics (higher score means higher ability) and their average perception of maths lesson difficulty (higher score means harder). The results from correlation analysis are summarised in Table 9.

Table 9: Correlations between teachers' reported pedagogy and average students' scores

Correlations

		TeacherScore	Average StudentScore	SD	Min	Max	AverageAbility	AverLessDiff
TeacherScore	Pearson Correlation	1.000	.178**	.171**	-.005	.207**	-.148*	-.052
	Sig. (2-tailed)		.007	.009	.938	.002	.025	.432
	N	230.000	230	230	230	230	230	230
AverageStudentScore	Pearson Correlation	.178**	1.000	.253**	.561**	.750**	.156**	-.202**
	Sig. (2-tailed)	.007		.000	.000	.000	.000	.000
	N	230	752.000	752	752	752	752	752
SD	Pearson Correlation	.171**	.253**	1.000	-.004	.248**	-.019	.059
	Sig. (2-tailed)	.009	.000		.916	.000	.600	.108
	N	230	752	752.000	752	752	752	752
Min	Pearson Correlation	-.005	.561**	-.004	1.000	.093*	.074*	-.104**
	Sig. (2-tailed)	.938	.000	.916		.010	.041	.004
	N	230	752	752	762.000	762	762	762
Max	Pearson Correlation	.207**	.750**	.248**	.093*	1.000	.042	-.328**
	Sig. (2-tailed)	.002	.000	.000	.010		.244	.000
	N	230	752	752	762	762.000	762	762
AverageAbility	Pearson Correlation	-.148*	.156**	-.019	.074*	.042	1.000	-.009
	Sig. (2-tailed)	.025	.000	.600	.041	.244		.807
	N	230	752	752	762	762	762.000	762
AverLessDiff	Pearson Correlation	-.052	-.202**	.059	-.104**	-.328**	-.009	1.000
	Sig. (2-tailed)	.432	.000	.108	.004	.000	.807	
	N	230	752	752	762	762	762	762.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The main observations that can be made from the above results are listed below:

- Teachers' scores seem to agree at some level with the average student scores (as of the positive statistically significant correlation).
- Association is stronger with the maximum student scores (positive and significant) and non-existent with the minimum scores (i.e. when more non-transmissionist practices were perceived).
- There are also some associations between the measures of pedagogy and students' perceived math ability as well as their perception of the lesson's difficulty. These relationships are better to be explored with the student-level dataset in the next section.

As a conclusion of this section we had demonstrated some agreement between the scores derived from students and their teachers in regards to their perception of ability. The weak/low correlations may be due to distributional properties of the scales (see Figure 10), the hierarchical structure of the datasets or even differences in perceptions (which will need to be explored qualitatively). Another point to be made based on the distributions is the consistent skewness towards transmissionist teaching.

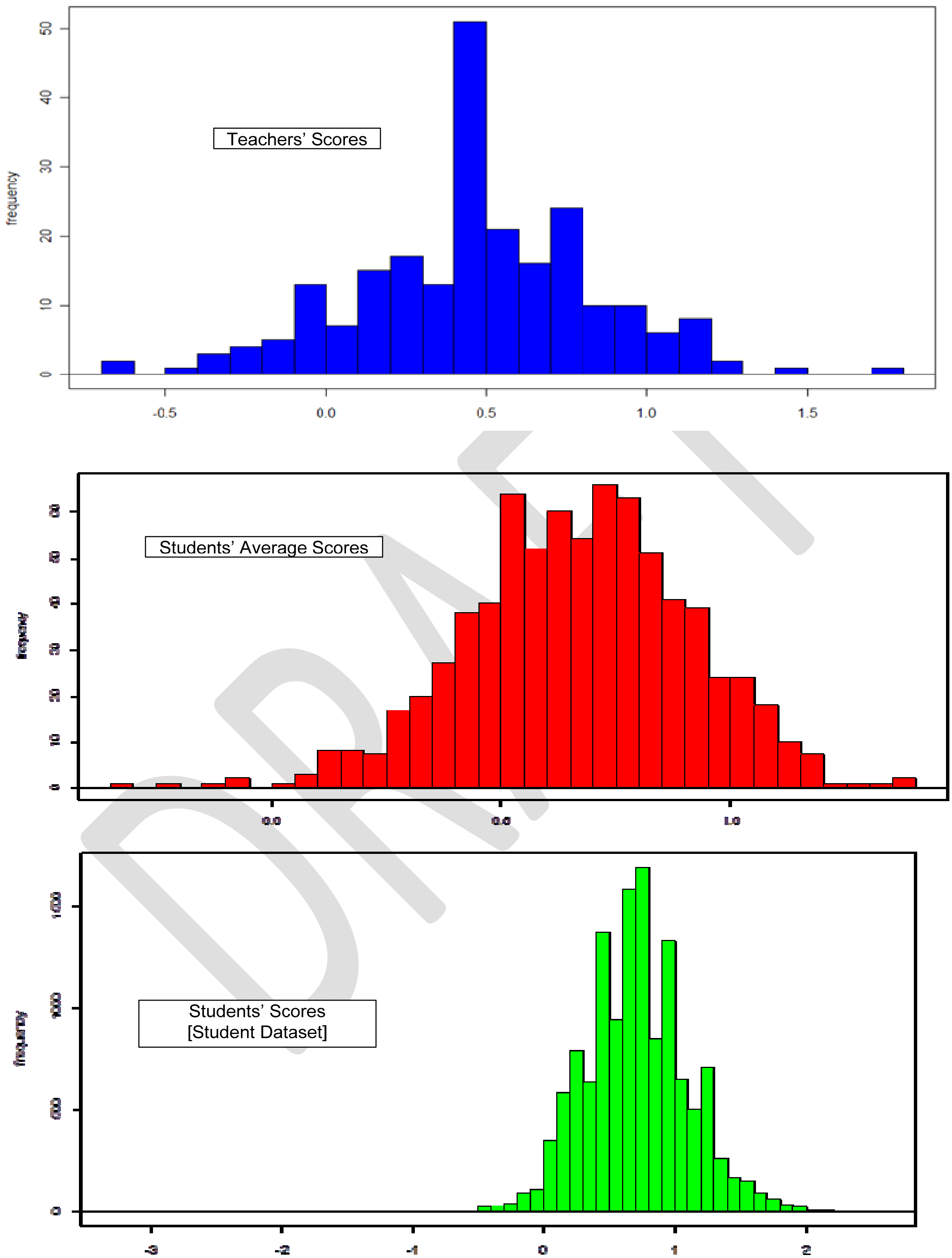


Figure 10: Distribution of the final measures (Histograms)

7. Further Statistical Modelling

Having established the validity of the constructed measures, and a certain degree for their agreement, the next crucial question is what they tell us about current practice in secondary mathematics education. Answer to this question will be more insightful once we complete data collection and also have linked longitudinal datasets to see changes over time. However, a preliminary descriptive mainly analysis is presented here based on the student dataset at DP1. In particular we seek to provide some evidence on the following:

- How teaching is changing across Year groups in secondary school?
- How is students' perception of teaching related with other variables of interest (gender, students' perception of their ability, students' favourite topics, etc.)?

As is evident from the definition of the questions above, we limit this descriptive analysis to the students' perception of pedagogy, to avoid complications and pitfalls because of the multilevel structure of the data in regards to teachers' scores.

Some preliminary findings are presented below, starting with Year group differences:

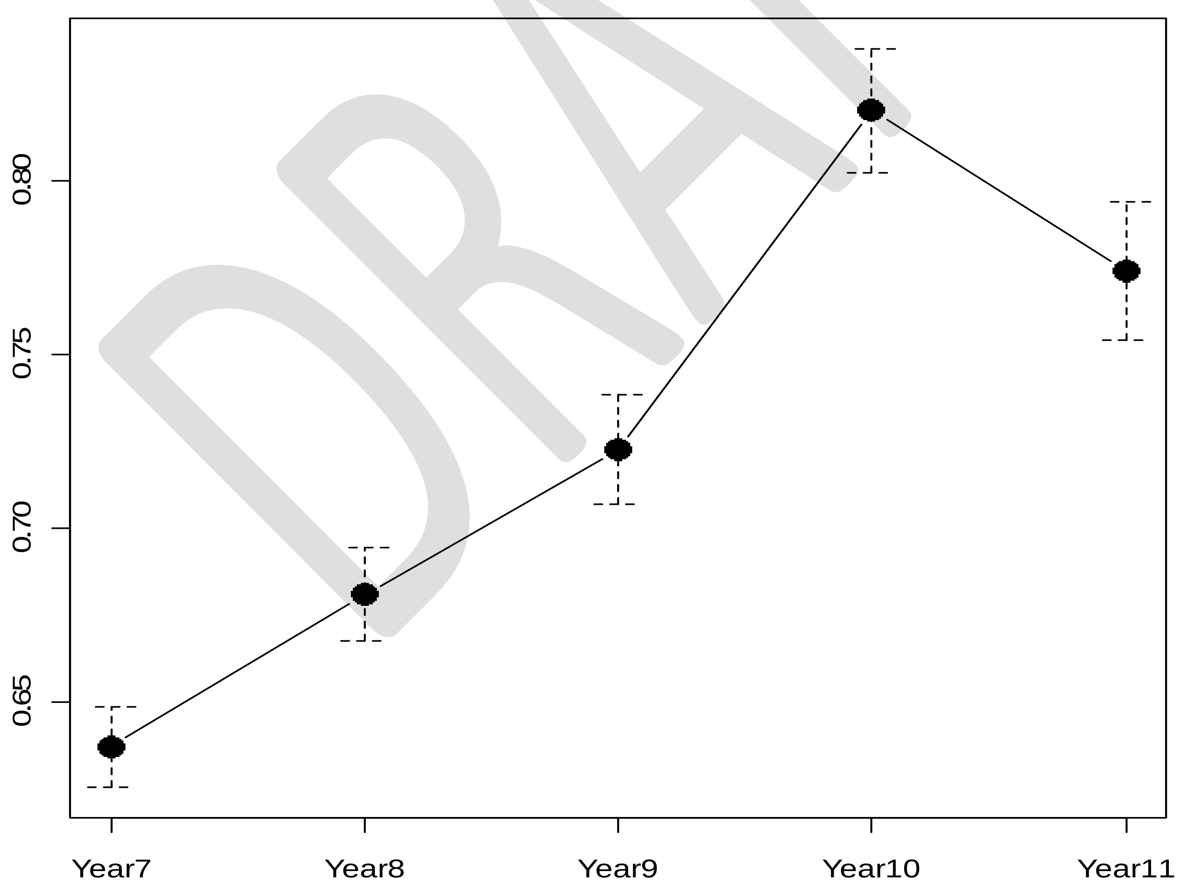


Figure 11: Plot of means of students' perception of pedagogy by Year Group

Figure 11 shows how the means of students' scores in the measure of their perception about transmissionist teaching are increasing as students move forward in Secondary school. The jump in their scores from Year 9 to 10 is considerably bigger than the previous years' increases, and there is also a noticeable decrease from Year 10 to 11 – however the mean for year 11 is still higher than that of the KS3 year groups. The differences even though small in magnitude, seem to be statistically significant as indicated by the confidence intervals around the mean scores. A preliminary conclusion then based on this observation: according to students' reported scores the teaching of mathematics seems to be increasingly transmissionist as students move from Year 7 throughout secondary schools, reaching its 'top' during the GCSE years.

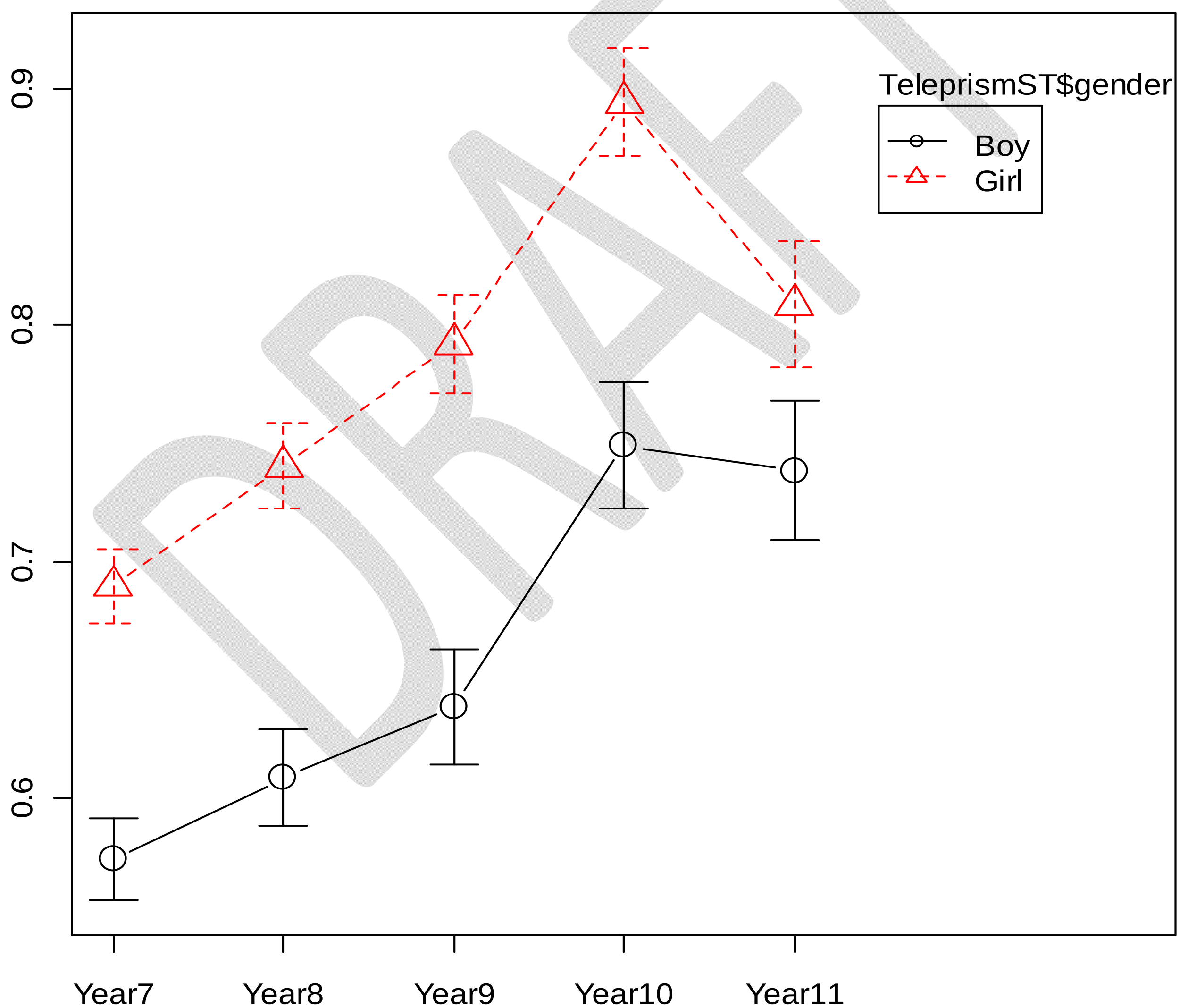


Figure 12: Plot of means of students' perception of pedagogy by Gender and Year Group

Figure 12, adds to the previous observation the consistently higher scores of pedagogy reported by girls independently of the Year group. The gender differences are much noticeable than most of the between Year groups' differences: for KS3 Year groups the differences in the boys weaken. The gender differences and their interaction with year group need further exploration. A fundamental question to follow this further is whether this picture remains the same when we control for classroom or the differences are due to classroom differences.

A final general question to explore is the association of this measure of pedagogy perception with variables relevant to students' mathematics dispositions. To respond to this we choose for this analysis two questions:

- A Likert type statement (we call Statement 19 for now) from Strongly Disagree to Strongly Agree (1 to 5): "I look forward to studying more mathematics in the future". Assuming an increasing maths disposition with the increase in students' ratings in this item, we correlated this with their scores of perceive pedagogy and found a significant negative correlation (Pearson r= -0.11, p-value < 2.2e-16, df=12995). This suggests that the more the students rate the teaching they experience as more traditional/transmissionist the less disposed they appear to continue studying mathematics.
- A second indicator of dispositions towards mathematics was derived from the open ended answers to two questions: What is your favourite and least favourite topic in school? Since our interest is mathematics here we constructed a new variable to denote whether the student reported mathematics either as their (a) favourite topic, (b) least favourite (worst) or (c) they did not include maths in their answer (indifferent). This categorical variable was then associated with the variable of interest here. As shown with Figure 13 there is a noticeable difference in the means of the students' groups: the mean perception of pedagogy for students who reported maths as their favourite topic is considerably lower than the group of students for whom maths is the least favourite subjects of who were indifferent about maths. A tentative conclusion again: students engage more with mathematics in less transmissionist learning environments.

A final point to be made in response to these preliminary results is that they are still preliminary and we don't want to make claims that any kind of teaching is bad, as given by negative associations with variables more relevant to mathematics engagement. We also found evidence that higher scores in students' perceptions of pedagogy are also associated positively with other variables such as their intention to go to university.

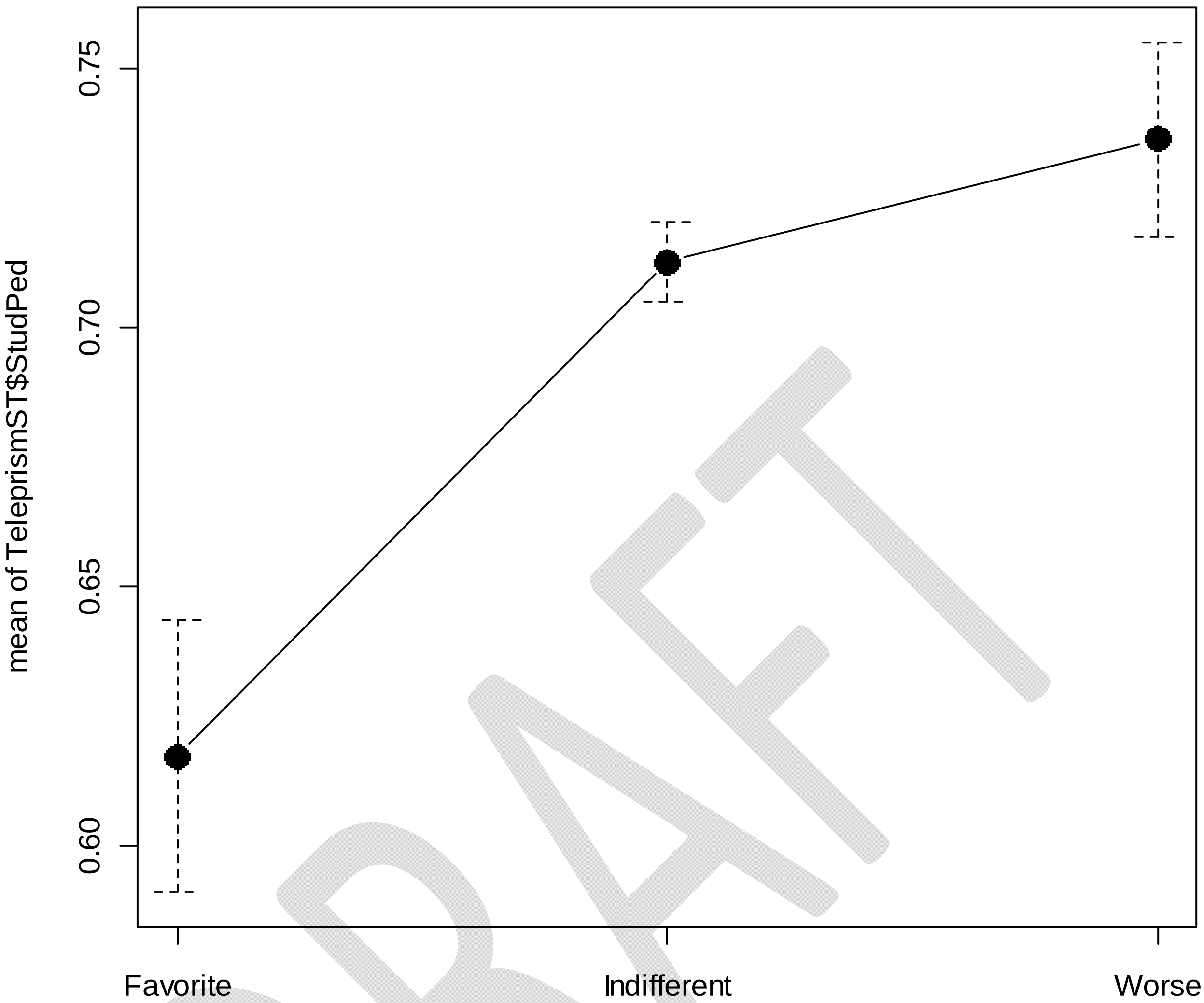


Figure 13: Plot of means of students' perception of pedagogy by mathematics preference

Limitations & Further work

In reading through this work the following should be considered:

- This is a tentative preliminary analysis
- Validation should and will be complemented with qualitative data from interviews with students and teachers (for a preliminary qualitative analysis of students interviews, see Qasim et al., BERA 2012)
- Some of the resulting associations may be masked by interactions with other variables. These will be further tested with generalised linear models.
- Multilevel modelling will also be pursuit to deal with the hierarchical structure of the data.

Discussion & Concluding Points

In summary, the paper shows:

- How we developed and validated two measures of what we call 'students' and teachers' perception of transmissionist teaching', respectively in secondary mathematics. Both measures could be considered robust in their current state, however due to the preliminary nature of this analysis we keep our reservations for more validity checks to ensure invariability across different groups and to account for the hierarchical structure of the data.
- Validation results also gave some interesting insights into the detail of teaching practices: for example we found that formal presentations and extended investigations become a less frequent practice as students move upwards in Secondary school (i.e. based on differences in items' functioning between Year groups).
- The distribution of both teachers and students in these two instruments seems to be skewed towards the higher ends of the scales, which correspond to more transmissionist/traditional teaching practices. As a tentative conclusion based on this we could argue that teaching secondary mathematics is currently perceived to be highly transmissionist from both students' and teachers' perspectives.
- Students and teachers perceptions about the same (assuming) class-cases of pedagogy appeared to be (weakly) positively correlated. The statistical significance gives some evidence of the hypothesis of agreement between these scores.
- Further analysis of the students' data indicated an increasing trend of their scores of transmissionist pedagogy as students move from Year 7 throughout secondary schools. Gender differences were also evident with girls consistently reporting experiences of more transmissionist teaching than boys.
- Finally we found some associations regarding students' engagement in mathematics and transmissionist teaching: One suggests that students tend to be less disposed to continue studying mathematics in more transmissionist classes and the other supports the claim that students engage more with mathematics in less transmissionist learning environments.

A final point should be emphasised again: these results are preliminary, tentative and not exhaustive. We have not yet explored how teaching is related to students' actual attainment and future aspirations. As already mention an initial finding in this area suggests an association of transmissionist teaching with students' intention to go to university. Moreover, in no way we judge the quality of this teaching. This is just the beginning of a story, mapping the general scene.

References

ACME. (2009). The Mathematics Education Landscape in 2009. *A report of the Advisory Committee on Mathematics Education (ACME) for the DCSF/DIUS STEM High Level Strategy Group Meeting (12 June 2009) retrieved from* http://www.acme-uk.org/downloaddoc.asp?id=139 *[March, 2010].*

Aldridge, J. M., Fraser, B. J., Taylor, P. C., & Chen, C.-C. (2000). Constructivist learning environments in a crossnational study in Taiwan and Australia. *International Journal of Science Education, 22*(1), 37-55.

Allexsaht-Snider, M., & Hart, L. E. (2001). "Mathematics for All": How Do We Get There? *Theory Into Practice, 40*(2), 93-101.

Archer, L., Halsall, A., Hollingworth, S., & Mendick, H. (2005). *'Dropping out and drifting away': an investigation of factors affecting inner-city pupils' identifies, aspirations and post-16 routes*. London: Institute for Policy Studies in Education.

Askew, M., Brown, M., Rhodes, V., Johnson, D., & Wiliam, D. (1997). *Effective Teachers of Numeracy (Final Report)*. London: King's College.

Bandura, A. (1978). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Adv. Behav. Res. Ther., 1*, 139-161.

Bandura, A., & Locke, E. A. (2003). Negative Self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87-99.

Bell, A. (1993). Principles for the Design of Teaching. *Educational Studies in Mathematics, 24*(1), 5-34.

Black, P., & Wiliam, D. (1998). *Inside the black box: raising standards through classroom assessment*. London: King's College.

Boaler, J. (2002). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching*. New Jersey: Lawrence Erlbaum Associates.

Boaler, J., & Greeno, J. (2000). Identity, Agency and Knowing in Mathematics Worlds. In J. Boaler (Ed.), *Multiple Perspectives on Mathematics Teaching and Learning*. Westport: Ablex Publishing.

Bohlig, M., Fisher, W. P. J., Masters, G. N., & Bond, T. (1998). Content Validity and Misfitting Items. *Rasch Measurement Transactions, 12*(1), 607.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. NJ: Lawrence Erlbaum Associates Inc.

Bong, M. (2001). Predictive utility of subject-, task-, and problem-specific self-efficacy judgments for immediate and delayed academic performances. *Journal of Experimental Education, 70*(2), 133-162.

Bong, M. (2004). Academic motivation in self-efficacy, task value, achievement goal orientations, and attributional beliefs. *Journal of Educational Research, 97*(6), 287-297.

Bowles, R. (2003). Rejecting Best Items? *Rasch Measurement Transactions, 17*(1), 917.

Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology, 46*(2), 129-149.

Cao, Z., Bishop, A., & Forgasz, H. (2007). Perceived parental influence on mathematics learning: a comparison among students in China and Australia. *Educational Studies in Mathematics, 64*(1), 85-106.

Cotton, T. (2001). Mathematics teaching in the real world. In P. Gates (Ed.), *Issues in mathematics teaching* (pp. 23-37). London: RoutledgeFalmer.

Cuban, L. (1983). How did teachers teach, 1890–1980. *Theory Into Practice, 22*(3), 160-165.

de Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical Evidence of a Comprehensive Model of School Effectiveness: A Multilevel Study in Mathematics in the 1st Year of Junior General Education in The Netherlands. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 15*(1), 3 - 31.

Driessen, G., & Sleegers, P. (2000). Consistency of Teaching Approach and Student Achievement: An Empirical Test. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 11*(1), 57 - 79.

Ernest, P. (1991). *The philosophy of mathematics education*. Basingstoke: Falmer.

Fraser, B. J. (1998). Classroom Environment Instruments: Development, Validity and Applications. *Learning Environments Research, 1*, 7-33.

Gwilliam, L. R., & Betz, N. E. (2001). Validity of Measures of Math- and Science-Related Self-Efficacy for African Americans and European Americans. *Journal of Career Assessment, 9*(3), 261-281.

Harwood, W. S., Hansen, J., & Lotter, C. (2006). Measuring teacher beliefs about inquiry: The development of a blended qualitative/quantitative instrument. *Journal of Science Education and Technology, 15*(1), 69-79.

Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., Miu-Ying Chui, A., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study*. US Department of Education: NCES: National Center for Educational Statistics.

Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. . *Journal of Higher Education 65*(1), 58-74.

Lapointe, J. M., Legault, F., & Batiste, S. J. (2005). Teacher interpersonal behavior and adolescents' motivation in mathematics: A comparison of learning disabled, average, and talented students. *International Journal of Educational Research, 43*(1-2), 39-54.

Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness *Journal of Applied Measurement 3*(1), 85-106.

Lopez, W. A. (1995). Rating scales and shared meaning. *Rasch Measurement Transactions, 9*(2), 434.

Lopez, W. A. (1996). Communication Validity and Rating Scales. *Rasch Measurement Transactions, 10* (1), 482-483.

Marat, D. (2005). Assessing mathematics self-efficacy of diverse students from secondary schools in Auckland: Implications for academic achievement. *Issues in Educational Research, 15*(1), 37-68.

McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among Instructional Practices, Curriculum, and Student Achievement: The Case of Standards-Based High School Mathematics. *Journal for Research in Mathematics Education, 32*(5), 493-517.

Meece, J. L., Herman, P., & McCombs, B. L. (2003). Relations of learner-centered teaching practices to adolescents' achievement goals. *International Journal of Educational Research, 39*, 457-475.

Mendick, H. (2005). Only connect: troubling oppositions in gender and mathematics. *International Journal of Inclusive Education 9*(2), 161-180.

Messick, S. (1988). The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33-45). London: Lawrence Erlbaum Associates, Publishers.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Third ed., pp. 13-103). USA: American Council of Education and the Oryx Press.

NCES. (2000). *Mathematics and Science in the Eighth Grade: Findings from the Third International Mathematics and Science Study*: US Department of Education Office of Educational Research and Improvement.

Nielsen, I. L., & Moore, K. A. (2003). Psychometric data on the mathematics self-efficacy scale. *Educational and Psychological Measurement, 63*(1), 128-138.
O'Brien, V., Martinez-Pons, M., & Kopala, M. (1999). Mathematics self-efficacy, ethnic identity, gender, and career interests related to mathematics and science. *Journal of Educational Research, 92*(4), 231-235.
OFSTED. (2006). *Evaluating mathematics provision for 14–19-year-olds.*
OFSTED. (2008). Mathematics - Understanding the Score.
Pampaka, M., Kleanthous, I., Hutcheson, G. D., & Wake, G. (2011). Measuring mathematics self-efficacy as a learning outcome. *Research in Mathematics Education, 13*(2), 169-190.
Pampaka, M., Williams, J., & Hutchenson, G. (2011). Measuring students' transition into university and its association with learning outcomes. *British Educational Research Journal*, First published on: 14 September 2011 (iFirst).
Pampaka, M., & Williams, J. S. (2010). *Measuring Mathematics Self Efficacy of students at the beginning of their Higher Education Studies.* Paper presented at the Proceedings of the British Congress for Mathematics Education (BCME) (pp. 159-166).
Pampaka, M., Williams, J. S., Davis, P., & Wake, G. (2008). *Measuring pedagogic practice: a measure of 'teacher-centrism'.* Paper presented at the The American Educational Research Association annual conference (AERA).
Pampaka, M., Williams, J. S., Hutchenson, G., Black, L., Davis , P., Hernandez-Martines, P., & Wake, G. (under review). Measuring Alternative Learning Outcomes: Dispositions to Study in Higher Education. *Journal of Applied Measurement.*
Pampaka, M., Williams, J. S., Hutcheson, G., Wake, G., Black, L., Davis, P., & Hernandez - Martinez, P. (2012). The association between mathematics pedagogy and learners' dispositions for university study. *British Educational Research Journal, 38*(3), 473-496.
Pietsch, J., Walker, R., & Chapman, E. (2003). The Relationship Among Self-Concept, Self-Efficacy, and Performance in Mathematics During Secondary School. *Journal of Educational Psychology, 95*(3), 589-603.
Pustjens, H., Van de gaer, E., Van Damme, J., & Onghena, P. (2004). Effect of Secondary Schools on Academic Choices and on Success in Higher Education. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 15*(3), 281 - 311.
Raudenbush, S. W., Johnson, C., & Sampson, R. J. (2003). A Multivariate, Multilevel Rasch Model with Application to Self-Reported Criminal Behavior. *Sociological Methodology, 33*, 169-211.
Roberts, G. (2002). *SET for Success. The supply of people with science, technology, engineering and mathematics skills.* London: HM Stationery Office.
Roelofs, E., Visser, J., & Terwel, J. (2003). Preferences for various learning environments: Teachers' and parents' perceptions. *Learning Environments Research, 6*(1), 77-110.
Ryan, J. T., & Williams, J. S. (2007). *Children's mathematics 4 - 15.* Milton Keynes: Open University Press.
Schuh, K. L. (2004). Learner-centered principles in teacher-centered practices? *Teaching and Teacher Education, 20*(8), 833-846.
Shuell, T. J. (1993). Toward an Integrated Theory of Teaching and Learning. *Educational Psychologist, 28*(4), 291 - 311.
Silva, F. (1993). *Psychometric Foundations and Behavioral Assessment.* London: Sage Publications.
Smith, A. (2004). *Making mathematics count – the report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education.* London: DfES.
Smith, R. M., Schumacker, R. E., & Busch, M. J. (1998). Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement, 2*(1), 66-78.

Swan, M. (2000). The Purpose of Mathematical activities and Pupils' Perceptions of them. *Research in Education, 63*, 199-223.
Swan, M. (2006). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research in Education*(75), 58-70.
Swanson, C. B., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP State Assessments. *Educational Evaluation and Policy Analysis, 24*(1), 1-27.
Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 67-114). London Lawrence Erlbaum Associates, Publishers.
van Langen, A., & Dekkers, H. (2005). Cross-national differences in participating in tertiary science, technology, engineering and mathematics education. *Comparative Education, 41*(3), 329-350.
Vermunt, J., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction, 9*(3), 257 - 280.
Wake, G., & Pampaka, M. (2008). The central role of the teacher - even in studen centred pedagogies. *Proceedings of the Joint Meeting of the 32nd Conference of the International Group for the Psychology of Mathematics Education and the XX North American Chapter. Morelia, Michoacán, México.*
Webster, B. J., & Fisher, D. L. (2003). School-level environment and student outcomes in mathematics. *Learning Environments Research, 6*(3), 309-326.
Westerhof, K. J. (1992). On the Effectiveness of Teaching: Direct Versus Indirect Instruction. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 3*(3), 204 - 215.
Wilkins, J. L. M., & Ma, X. (2003). Modeling Change in Student Attitude Toward and Beliefs about Mathematics. *The Journal of Educational Research, 97*(1), 52-63.
Wolf, A. (2000). Mathematics for some or mathematics for all? Curious UK practices in international context. In C. Tikly & A. Wolf (Eds.), *The maths we need now: demands, deficits and remedies* (pp. 104-137). London: University of London Institute of Education.
Wolf, A. (2002). *Does Education Matter? Myths About Education And Economic Growth.* London: Penguin Press.
Wolfe, E. W., & Smith Jr., E. V. (2007a). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part I - Instrument Development Tools. *Journal of Applied Measurement, 8*(1), 97-123.
Wolfe, E. W., & Smith Jr., E. V. (2007b). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part II - Validation Activities. *Journal of Applied Measurement, 8*(2), 204-234.
Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97-116.
Wright, B. D. (1994). Data Analysis and Fit. *Rasch Measurement Transactions, 7*(4), 324.
Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis.* Chicago: MESA Press.

APPENDIX: Frequency bars of different teaching practices happening during maths lessons

Key	Never	Rarely	Sometimes	Always

Item name	Frequency bars	Never	Rarely	Sometimes	Always
The teacher asks us questions.		126 (1%)	396 (3%)	3415 (26%)	9092 (70%)
The teacher tells us which questions/activities to do.		161 (1%)	403 (3%)	2964 (23%)	9377 (73%)
The teacher asks us to explain how we get our answers.		142 (1%)	576 (4%)	4706 (36%)	7555 (58%)
We listen to the teacher talk about the topic.		221 (2%)	786 (6%)	5163 (40%)	6654 (52%)
The teacher expects us to remember important ideas learnt in the past.		265 (2%)	1103 (9%)	5867 (45%)	5699 (44%)
We copy the teacher's notes from the board.		322 (3%)	1544 (12%)	6411 (50%)	4539 (35%)
The teacher gives us problems to investigate.		436 (3%)	1568 (12%)	7383 (57%)	3530 (27%)
The teacher asks us what we already know about a lesson topic.		523 (4%)	1511 (12%)	6387 (50%)	4426 (34%)
We discuss ideas with the whole classroom.		611 (5%)	1649 (13%)	6309 (49%)	4206 (33%)
The teacher uses the computer to teach some topics.		620 (5%)	1882 (15%)	6757 (52%)	3682 (28%)
We talk with other students about how to solve problems.		627 (5%)	1828 (14%)	7329 (57%)	3009 (24%)
We work through exercises from the textbook.		966 (8%)	1885 (15%)	6007 (47%)	3911 (31%)
We use calculators.		736 (6%)	2388 (19%)	8205 (64%)	1479 (12%)
We ask other students to explain their ideas.		844 (7%)	2453 (19%)	6907 (54%)	2572 (20%)
We explain our work to the whole class.		966 (8%)	2495 (20%)	6488 (51%)	2739 (22%)
The teacher tells us to work more quickly.		844 (7%)	3344 (26%)	6472 (50%)	2279 (18%)
The teacher tells us what value the lesson topic has for future use.		1935 (15%)	3518 (27%)	5240 (41%)	2104 (16%)
We work together in groups on projects.		1707 (13%)	4035 (31%)	6375 (50%)	740 (6%)
What we learn is related with our out-of-school life.		2336 (18%)	3814 (30%)	5197 (41%)	1355 (11%)
We learn that mathematics is about inventing rules.		2800 (22%)	3929 (31%)	4791 (38%)	1133 (9%)
We get assignments to research topics on our own.		3626 (29%)	3918 (31%)	4255 (34%)	842 (7%)
The teacher starts new topics with problems about the world.		3572 (28%)	4205 (33%)	4058 (32%)	999 (8%)
We use computers.		2977 (23%)	4955 (39%)	4111 (32%)	743 (6%)
We do projects (assignments) that include other school subjects.		3597 (28%)	4544 (36%)	3825 (30%)	782 (6%)
We learn how mathematics has changed over time.		4163 (33%)	4271 (33%)	3496 (27%)	826 (6%)
We use other things like newspapers, magazines, or video.		6361 (50%)	4647 (36%)	1545 (12%)	200 (2%)