Measuring Pedagogies in Mathematics with the Rasch Model: from Secondary School to University and across countries

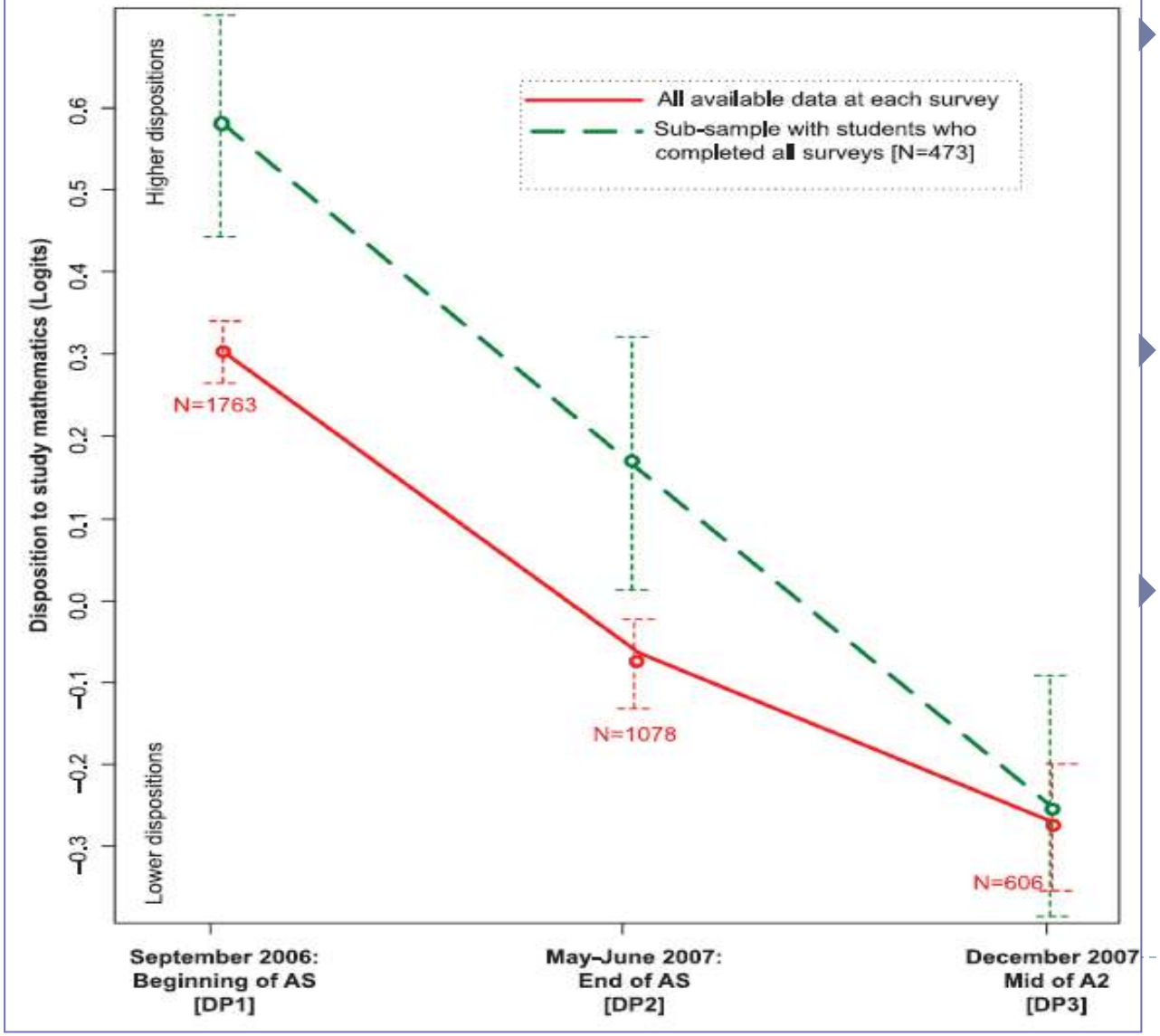
Maria Pampaka
The University of Manchester

6th UK Rasch User Group Meeting
20th March 2012

Outline

- Introduction – the STEM problem
- Background to the projects
- (college) Teachers' reported pedagogical practices (and effect on models of students' dispositions)
- Students' reported perception of pre-university pedagogical experience
- Comparison of measures from UK and Norway
- Some comparisons and associations
- (Pedagogy at Secondary school)
- Concluding remarks

Introduction: The STEM 'issue'



- STEM: Science Technology, Engineering and Mathematics
- Participation remains problematic
- Students dispositions are declining

Our focus

- Development of the measure of teacher self-reported pre-university pedagogy and its association with students' learning outcomes
- The 'conversion' of this pedagogy instrument into two measures of students' perceived pedagogical experience before and during their first year at university in UK and in Norway (cross-national comparisons)
- [The extension/development of these instruments backwards to capture secondary students' progression into secondary schools (Year 7 to 11)]

The Projects

- ESRC funded projects on transition to mathematically demanding subjects in UK Higher Education (HE): TransMaths
  - **TLRP: "Keeping open the door to mathematically demanding F&HE programmes" (2006 – 2008)**
  - **TransMaths: "Mathematics learning, identity and educational practice: the transition into Higher Education" (2008-2010)**
  - Lead PI: Prof Julian Williams

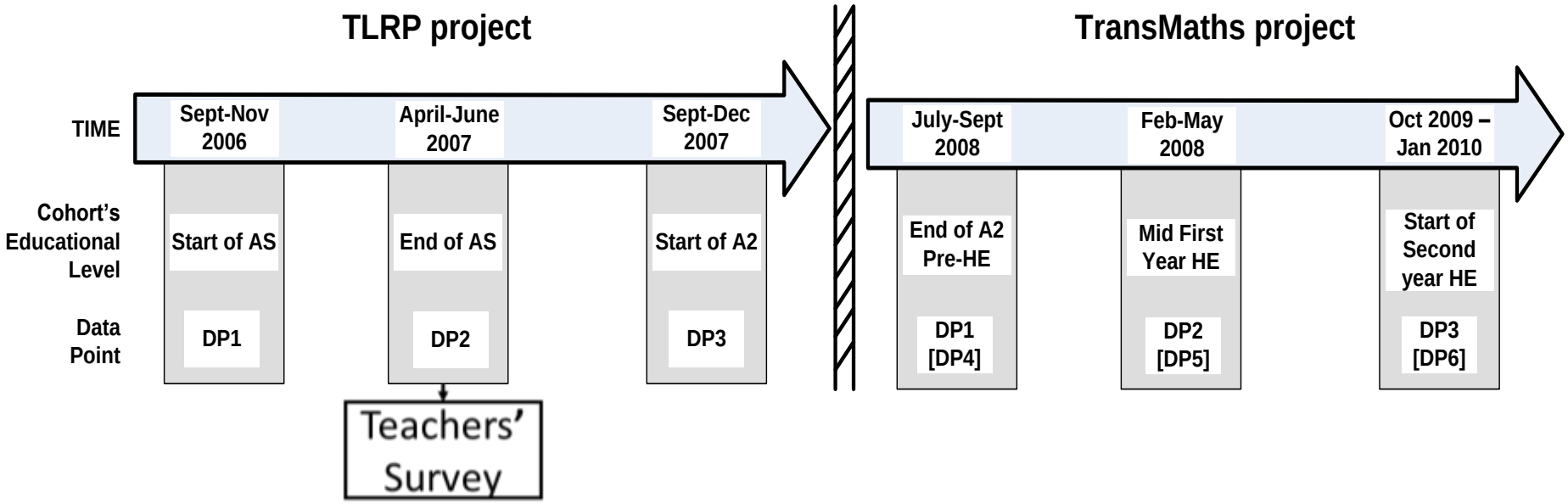
  www.transmaths.org

- An extension of this work in Norway: TransMaths-Norway
  - Lead PI: Prof Birgit Pepin

- Ongoing ESRC funded study of teaching and learning secondary mathematics in UK (2011-2014): Teleprism
  - PI: Dr Maria Pampaka

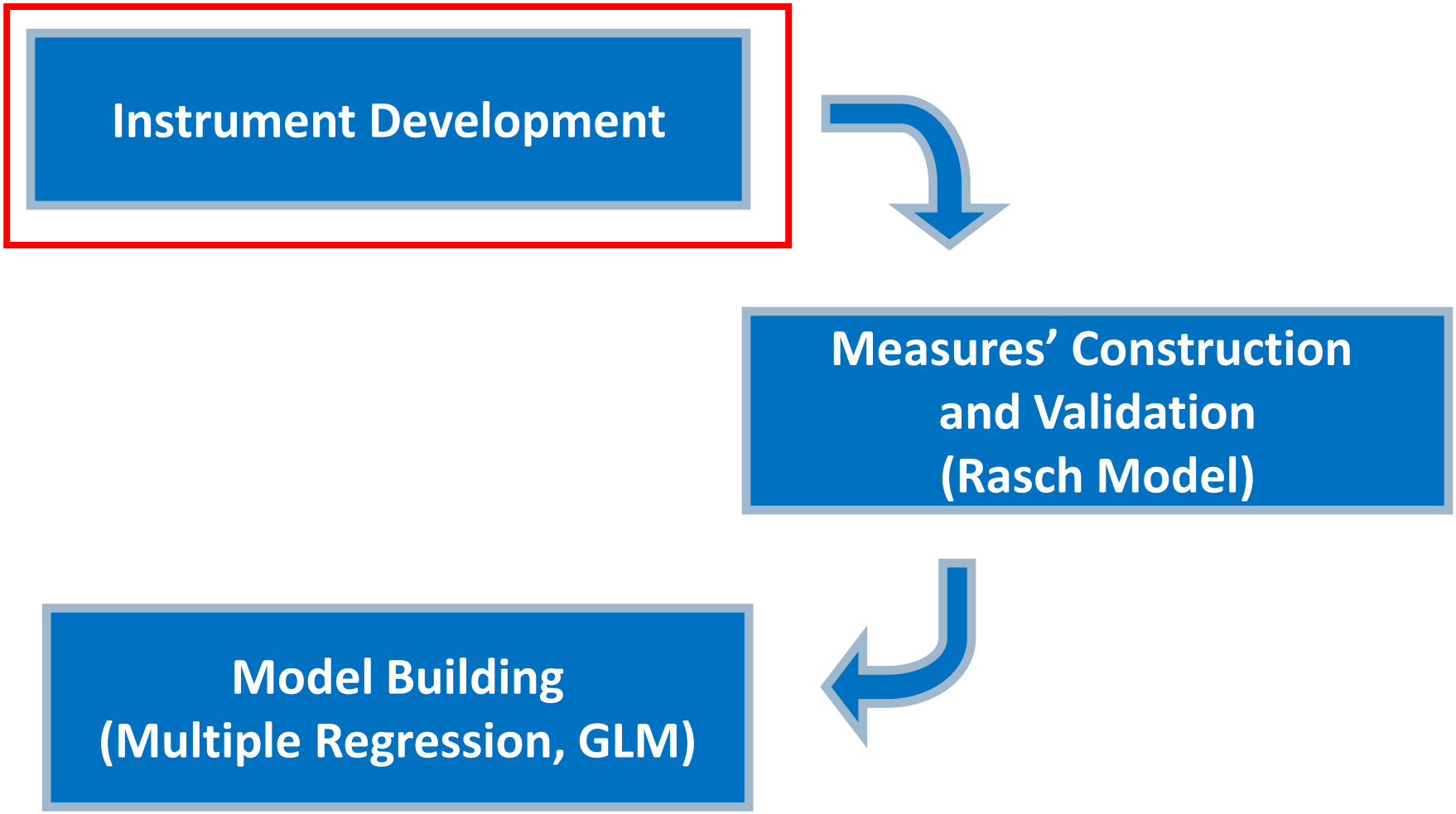
  www.teleprism.com

The TransMaths Project(s) Design

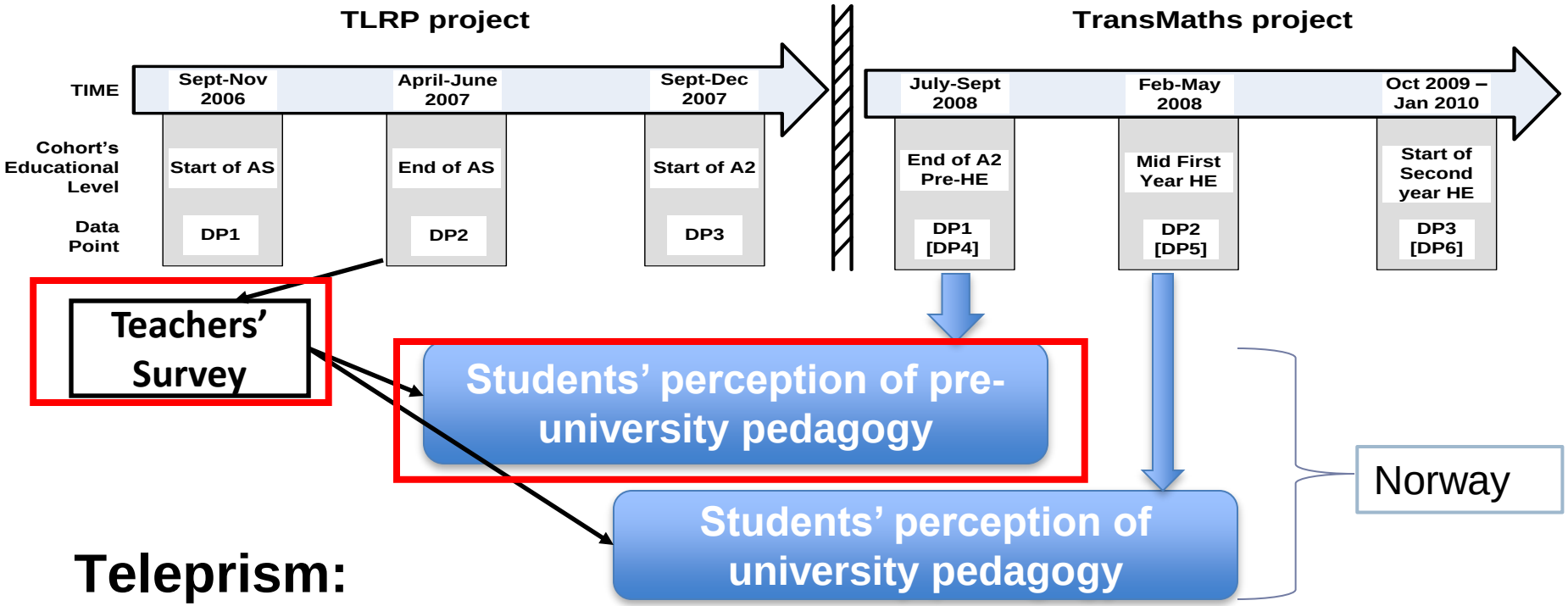


- TransMaths-Norway: University Transition in Norway
- TeLePriSM: Dispositions and Pedagogies at Secondary Mathematics UK (Year 7 to 11)

Analytical Framework



Instrumentation



Teleprism:

- Measures of teachers' self report teaching
- Students' perception of teaching
- Common items

Teacher Instrument Development

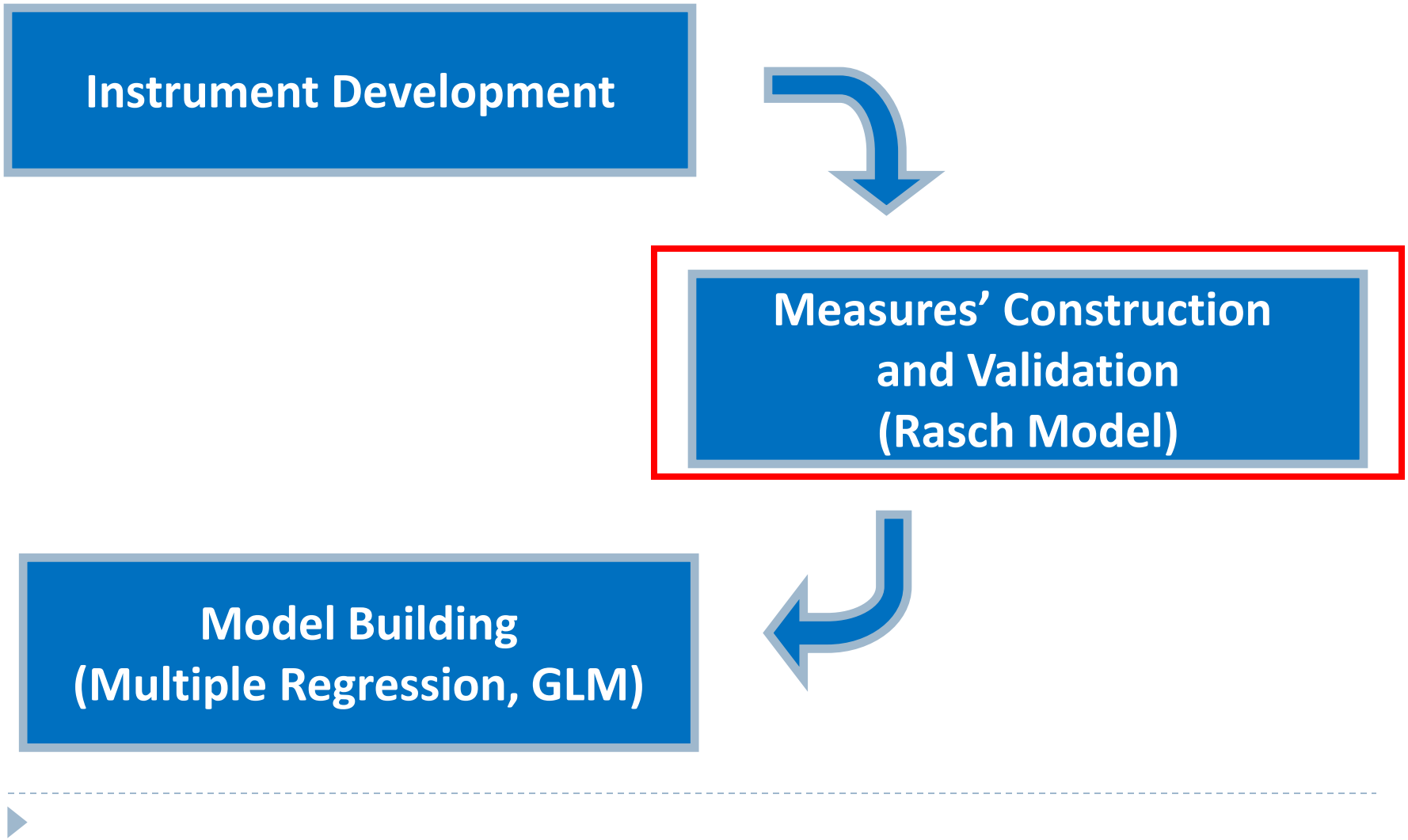
28 items

- 5 point Likert Scale (for frequency)
- Calibrated Swan's original data
- Re-calibration with 110 cases from TLRP project

SECTION B: ABOUT YOUR TEACHING OF THIS CLASS

	Tick one box in each row to show how often the following statements are true in your lessons with this class.	Almost never	Occasionally	About half the time	Most of the time	Almost always
1	Students work through exercises.					
2	Students work on their own, consulting a neighbour from time to time.					
3	Students use only the methods I teach them.					

Analytical Framework



Measurement Methodology

- ‘Theoretically’: Rasch Analysis
  - Partial Credit Model
  - Rating Scale Model (for the pedagogic measure)

- ‘In practice’ – the tools:
  - FACETS, Quest and Winsteps software

- Interpreting Results:
  - (Item) Fit Statistics (to ensure unidimensional measures)
  - Differential Item Functioning for ‘subject’ groups
  - Person-Item maps for hierarchy (person fit, as well)
  - Response category statistics for rating scale functionning

Constructing the measures – Validity [Unidimensionality - Fit]

B6: I encourage students to work more slowly

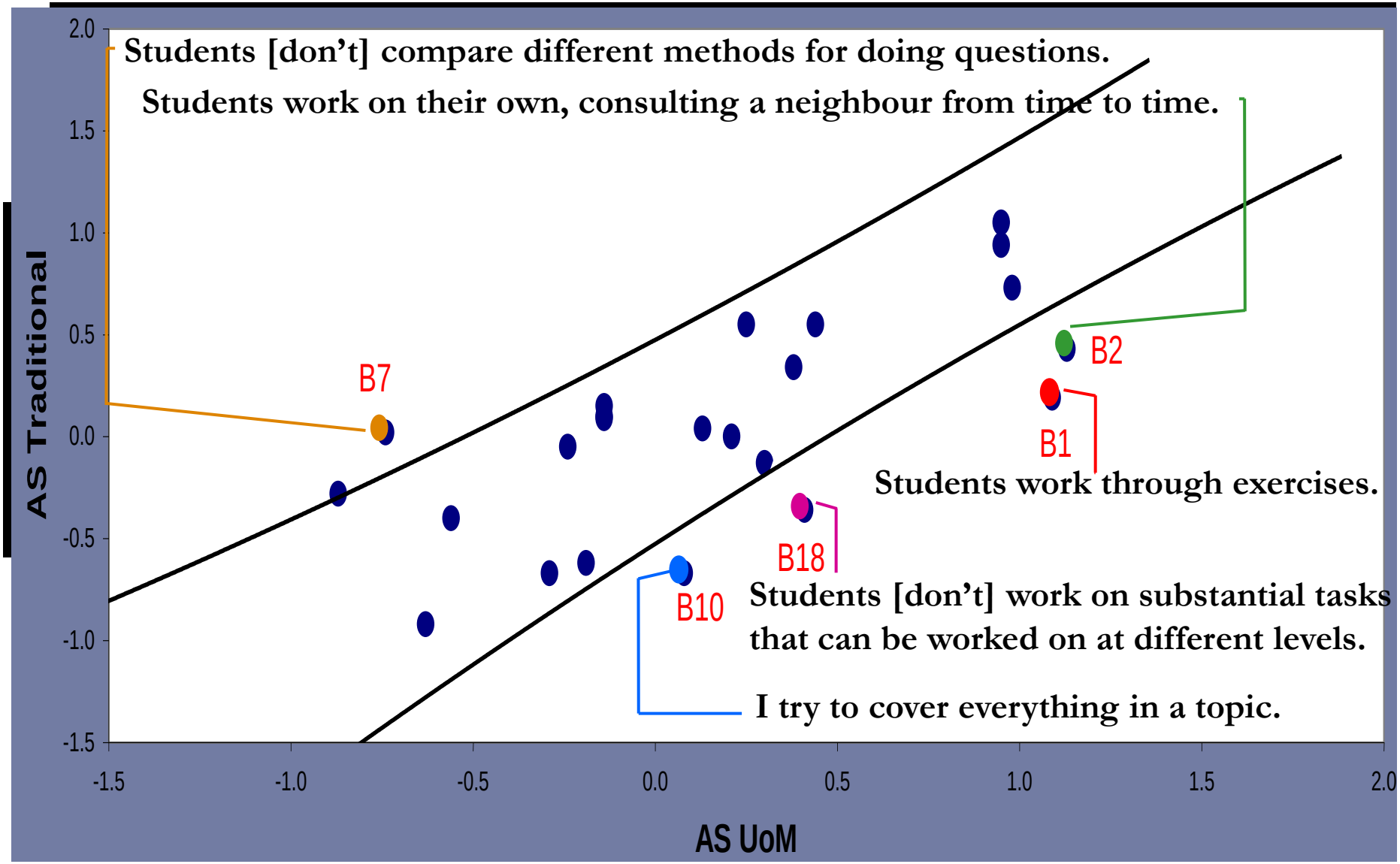
B24: I cover only the important ideas in a topic

Item	Entry No	Raw Score	Count	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
B1	1	355	111	.42	.10	0.8	-2	0.8	-1
B2	2	391	111	.05	.10	1.0	0	1.0	0
B4	4	382	111	.13	.10	0.9	-1	0.9	-1
B5	5	391	110	-.83	.13	1.0	0	0.9	0
B6	6	456	111	-.84	.13	1.5	2	1.5	2
B7	7	411	111	-.18	.11	0.7	-2	0.7	-2
B8	8	405	111	-.11	.11	1.4	2	1.3	2
B9	9	434	111	-.49	.12	0.5	-3	0.5	-3
B10	10	426	110	-.43	.12	1.1	0	1.1	0
B11	11	295	111	.96	.09	1.0	0	1.0	0
B12	12	391	111	.05	.10	0.7	-2	0.7	-2
B13	13	367	105	.07	.11	1.2	1	1.3	1
B14	14	268	110	1.18	.10	1.2	1	1.2	1
B15	15	302	111	.90	.09	0.7	-2	0.7	-2
B16	16	348	111	.48	.10	0.7	-2	0.8	-2
B17	17	460	109	-1.06	.14	0.8	-1	0.8	-1
B									1
B									0
B									-2
B									0
B22	22	396	111	-.03	.11	1.4	2	1.4	2
B23	23	382	109	.06	.11	1.2	1	1.2	1
B24	24	392	109	-.04	.11	1.5	3	1.5	3
B25	25	391	111	.05	.10	0.9	0	0.9	0
B26	26	444	109	-.77	.13	1.4	2	1.2	1
B28	27	359	110	.34	.10	1.1	0	1.1	0
Mean		386.5	110.2	.00	.11	1.0	-0.1	1.0	-0.1
SD		51.2	1.3	.58	.01	0.3	1.9	0.3	1.8

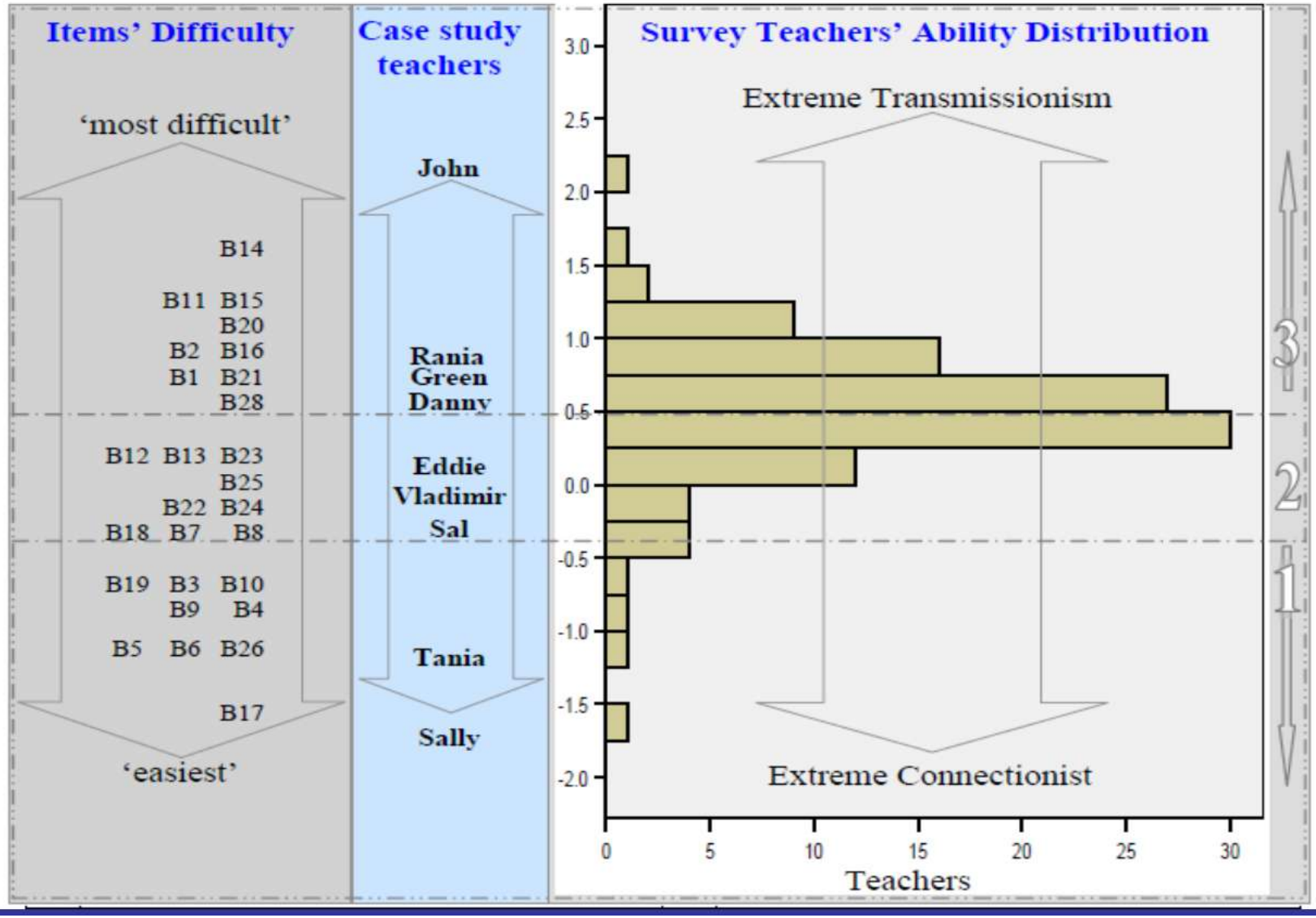
RMSE (Model) .11 Adj S.D. .57 Separation 5.13 Reliability .96
Fixed (all same) chi-square: 742.2 d.f.: 26 significance: .00
Random (normal) chi-square: 26.0 d.f.: 25 significance: .41

Constructing the measures [Validity across different groups (DIF)]

Use of Maths Vs AS Traditional Courses



A measure of "pedagogical style": "Teacher centricism" Scale



A measure of "pedagogical style": "Teacher centricism" Scale

I tend to follow the textbook closely

Students (don't) discuss their ideas

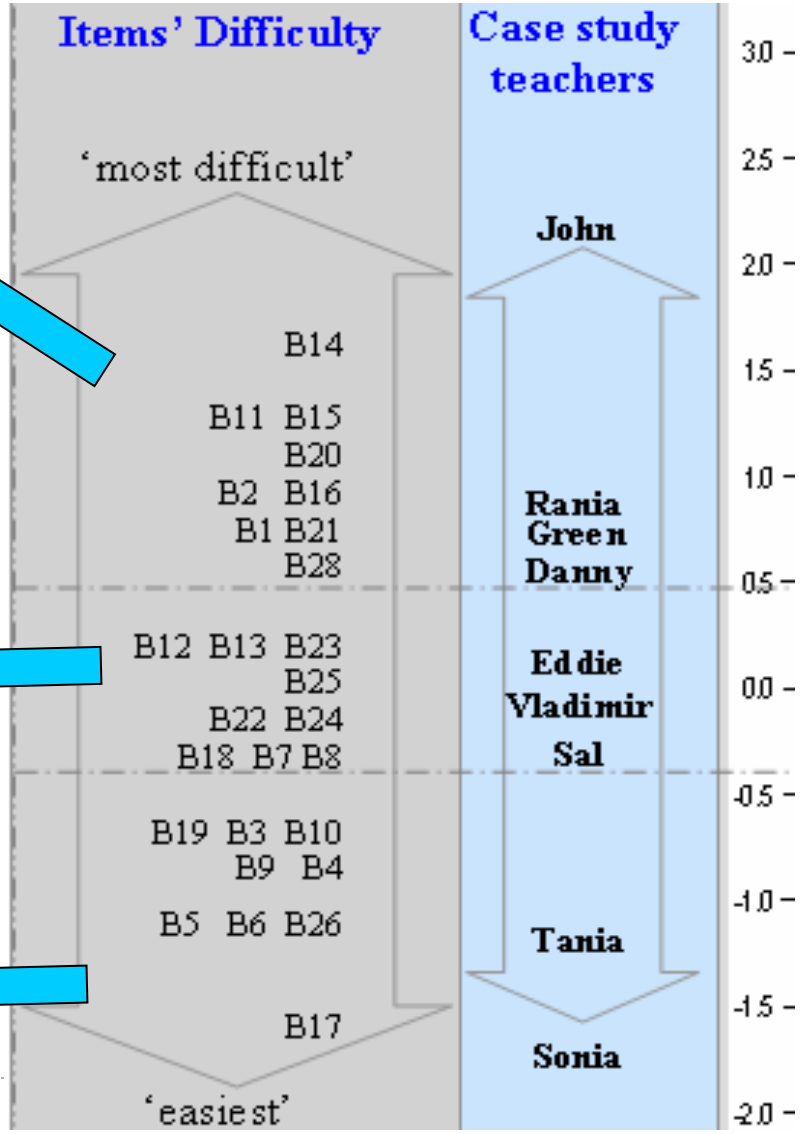
I encourage students to work more quickly

I teach each topic separately

I tell students which questions to tackle

I know exactly what maths the lesson will contain

Students (don't) invent their own methods

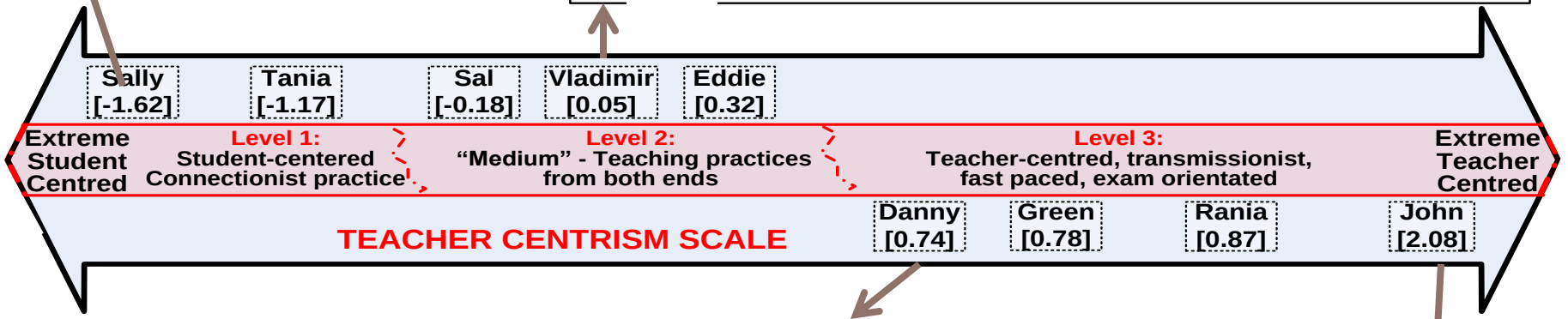


Validation supported with Qualitative Data

"… there's a sense that I've achieved the purpose…I've found out what they've come with and what they haven't come with so…we can work with that now"

"…. from the teachers that I've met and talked to… it seems to me that one of the big differences is, I mean I don't sort of use textbooks… [ ]…I want to get students to think about the math, I want students to understand, I want students to connect ideas together, to see all those things that go together and I don't think a text book did that…."

Here we work backwards, here the student has got a certain data and then trying to find a model for that so directly comes to their need. So they measure something, they take some reading and now they want to put a mathematical language to this finding and I always find a model for that language.

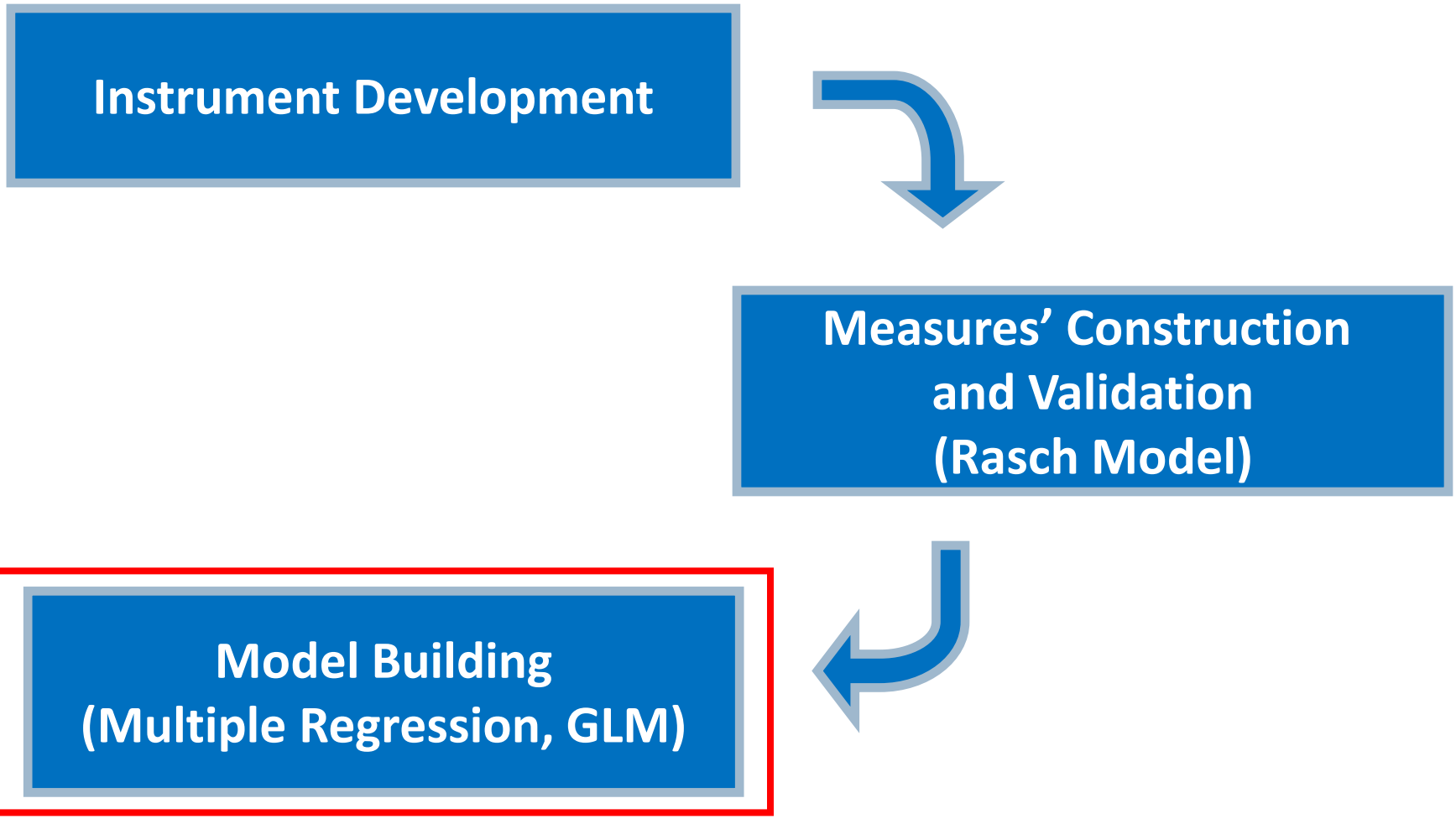


…. I think Powerpoint and text just switches everybody off if there is too much text on there. We don't want to read off the screen. That is more or less for me to do the explanations as we run through it and then consolidate that with them doing something.

"…I do tend to teach to the syllabus now…If it's not on I don't teach it. … but I do tend to say this is going to be on the exam, it's going to be worth X number of marks, that's why we're doing it."

"It's old fashion methods, there's a bit of input from me at the front and then I try to get them working, practising questions as quickly as possible, …"

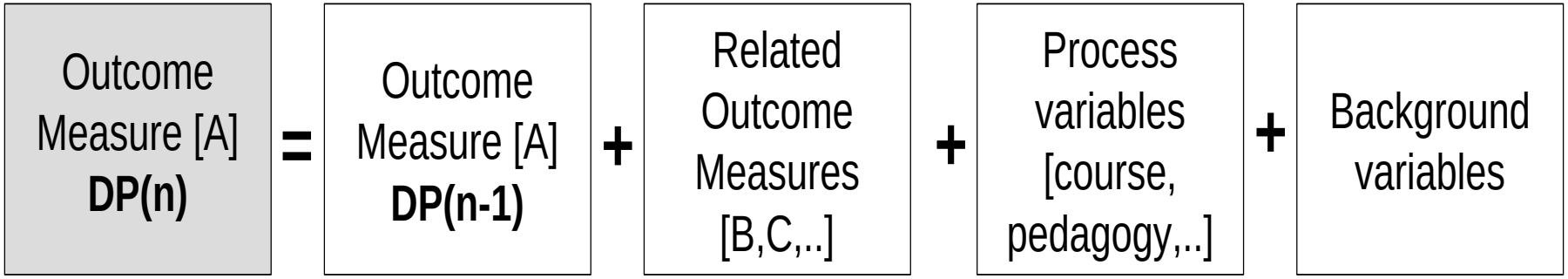
Using the measures to answer RQ



The TLRP sample

Data collection	Gender	Mathematics course		Total
		AS Trad	AS UoM	
DP1 – Beginning of AS	Male	773	340	1113
[September 2006]	Female	512	154	666
	Total DP1	**1285**	**494**	**1779 [1792]***
DP2 – End of AS	Male	428	236	664
[May–June 2007]	Female	302	110	412
	Total DP2	**730**	**346**	**1076 [1082]**
DP3 – Following year	Male	244	98	342
[November 2007–January 2008]	Female	215	47	262
	Total DP3	**459**	**145**	**604 [608]**

From Measurement to Modelling



Variables

- **Outcome of AS Maths (Grade, or Dropout)**
- **Background Variables**
- **Disposition Measures at each DP**
  - **Disposition to go into HE (HEdisp)**
  - **Disposition to study mathematically demanding subjects in HE (MHEdisp)**
  - **Maths Self Efficacy**
- **A score of 'pedagogy' based on teacher's survey**

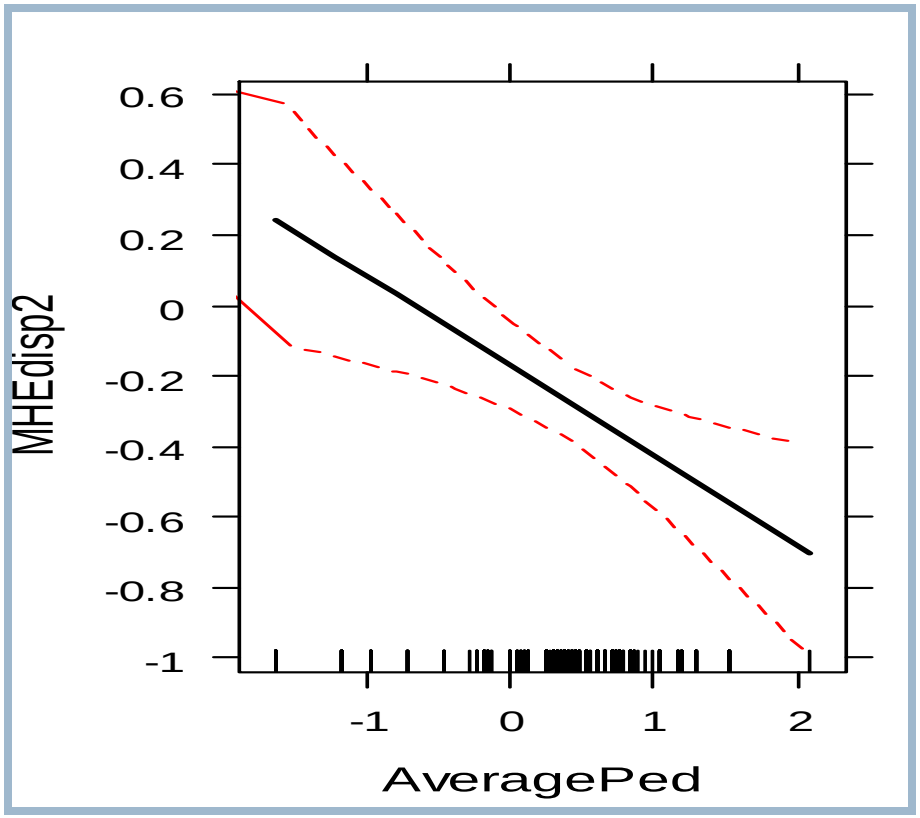
A model of HE Maths Disposition at the end of AS year

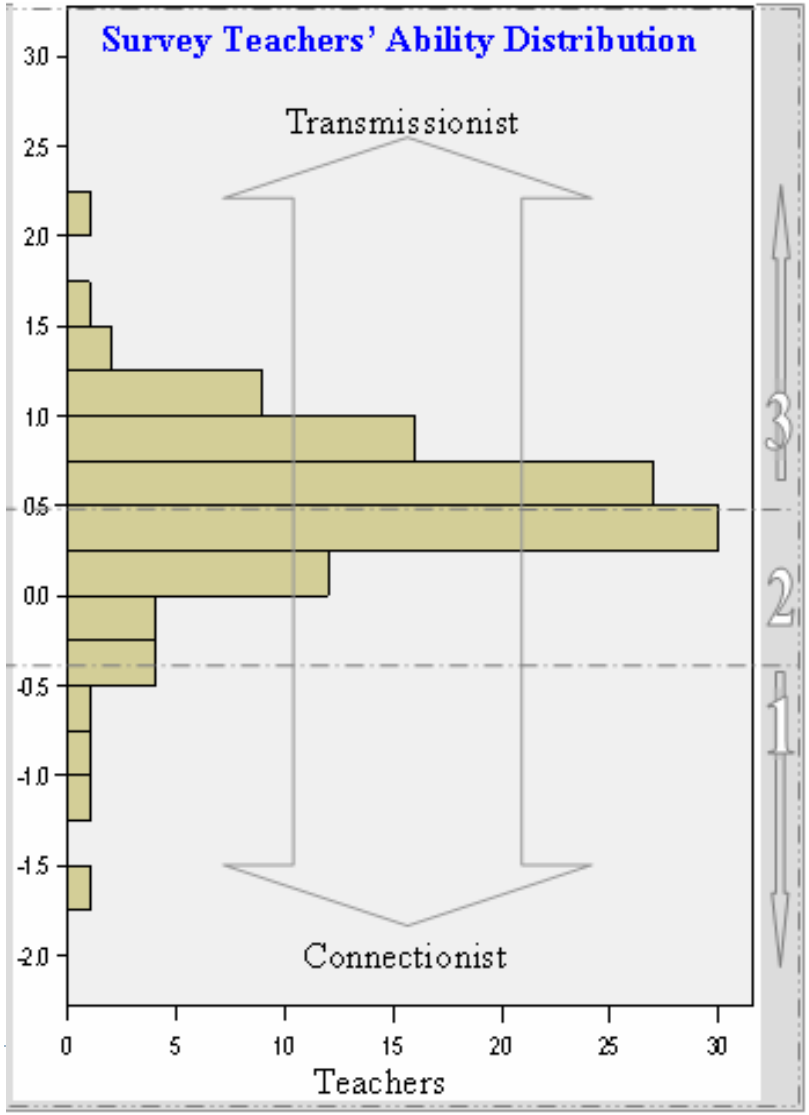
	Coefficient B	s.e.	t	p
(Constant)	–0.609	0.079	–7.754	3.11e-14
Maths disposition at DP1	0.645	0.035	18.342	<2e-16
Math demand of other subjects at DP2	0.191	0.031	6.179	1.09e-09
Pedagogical practice	–0.288	0.082	–3.519	0.0005

Notes: $F\ (3, 708) = 144, p < 0.001, R^2 = 0.379\ (Adjusted\ R^2 = 0.376)$.

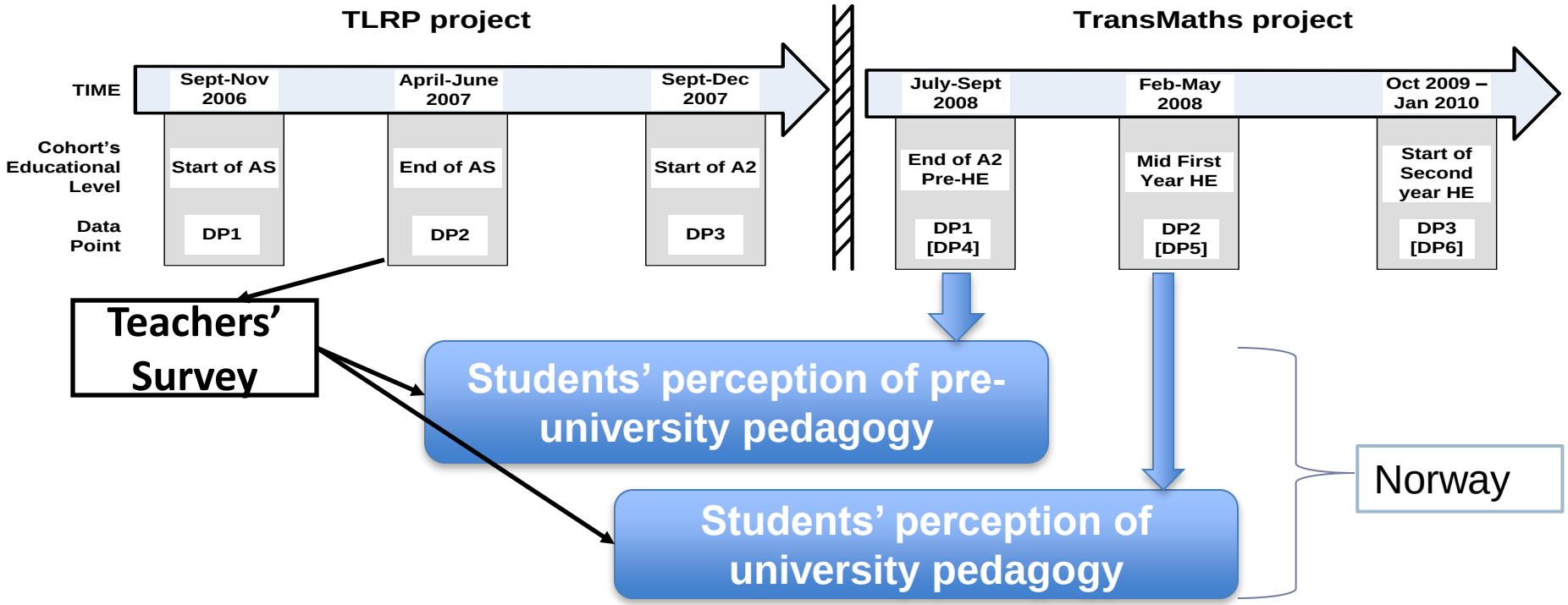
- Positive effect: Math Disposition at DP1, 'Mathematical demand of other subjects'
- Negative effect: pedagogy

Negative effect of Pedagogy





Extensions of this teacher instrument



A shortened version of the pedagogy instrument

Item description:	Measure	SE	Infit Mnsq	Outfit Mnsq
B3: Students use only the methods I teach them.	-0.67	0.14	0.9	0.9
B5: Students [don't] choose which questions they tackle.	-1.16	0.15	1.2	1
B7: Students [don't] compare different methods for doing questions.	-0.26	0.13	0.7	0.6
B11: I [don't] draw links between topics and move back and forth between topics.	1.35	0.11	1.2	1.2
B12: Students [don't] work collaboratively in small groups.	-0.01	0.13	0.8	0.8
B15: Students [don't] discuss their ideas.	1.19	0.11	0.8	0.8
B16: Students [don't] work collaboratively in pairs.	0.65	0.12	0.8	0.8
B17: Students [don't] invent their own methods.	-1.51	0.16	0.8	0.8
B19: I tell students which questions to tackle.	-0.64	0.14	1.3	1.3
B20: I encourage students to work more quickly.	1.06	0.11	1.4	1.4
B25: I teach each topic separately.	0	0.13	1.4	1.4
Person Summary Statistics: Separation 2.31 Reliability .84 **Item Summary Statistics**: Separation 6.92 Reliability .98 Category Statistics: OK				

Extensions: From TLRP teacher survey to a student instrument

[Circle 1, 2, 3 or 4, or tick the box if you don't know]

	Almost never	Some of the time	Most of the time	Almost always	DON'T KNOW
1. We (students) were using only the methods the teacher had taught us.	1	2	3	4	
2. We were choosing which questions to tackle.	1	2	3	4	
3. We were comparing different methods for doing questions.	1	2	3	4	
4. The teacher was drawing links between topics and moved back and forth between topics.	1	2	3	4	
5. We were working collaboratively in small groups.	1	2	3	4	
6. We (students) were discussing our ideas.	1	2	3	4	
7. We were working collaboratively in pairs.	1	2	3	4	
8. We were inventing our own methods.	1	2	3	4	
9. The teacher was telling us which questions to tackle.	1	2	3	4	
10. The teacher was encouraging us to work more quickly.	1	2	3	4	
11. The teacher was teaching each topic separately.	1	2	3	4	

Students' pre-uni pedagogical experience Item Fit Statistics (UK): N=1516 students

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	Nu	Items
4568	1499	3.0	3.11	-.50	.04	0.9	-4	0.9	-3	.32	1	item1
4125	1488	2.8	2.82	-.01	.03	0.9	-3	0.9	-3	.38	2	item2
3668	1494	2.5	2.47	.51	.03	0.8	-5	0.8	-5	.45	3	item3
3524	1478	2.4	2.39	.63	.03	1.0	0	1.0	0	.38	4	item4
4180	1493	2.8	2.85	-.06	.03	1.0	0	1.0	0	.48	5	item5
3750	1494	2.5	2.53	.42	.03	0.9	-4	0.8	-4	.55	6	item6
4150	1493	2.8	2.82	-.02	.03	1.1	1	1.1	2	.33	7	item7
4825	1486	3.2	3.31	-.90	.04	1.1	2	1.0	0	.44	8	item8
4195	1489	2.8	2.86	-.09	.03	1.0	0	1.0	0	.35	9	item9
3875	1488	2.6	2.63	.27	.03	1.4	9	1.5	9	.07	10	item10
4294	1475	2.9	2.96	-.25	.03	1.1	3	1.1	3	.24	11	item11
4104.9	1488.8	2.8	2.80	.00	.03	1.0	-0.3	1.0	-0.3	.36	Mean (Count: 11)	
368.3	6.8	0.2	0.26	.43	.00	0.2	4.1	0.2	4.0	.12	S.D.	

RMSE (Model) .03 Adj S.D. .43 Separation 12.53 Reliability .99
Fixed (all same) chi-square: 1634.3 d.f.: 10 significance: .00
Random (normal) chi-square: 10.0 d.f.: 9 significance: .35

Item 10: The teacher was encouraging us to work more quickly

Students' pre-uni pedagogical experience Item Fit Statistics (Norway): N=709

Obsvd Score	Obsvd Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	Nu	Items
2312	700	3.3	3.37	-.57	.06	1.0	0	1.1	1	.28	1	item1
2100	702	3.0	3.05	.07	.05	1.0	0	1.0	0	.35	2	item2
2110	695	3.0	3.10	-.02	.05	0.7	-5	0.7	-5	.46	3	item3
1954	679	2.9	2.93	.27	.05	0.8	-3	0.9	-2	.38	4	item4
2150	703	3.1	3.12	-.06	.05	1.1	2	1.1	1	.40	5	item5
2173	702	3.1	3.15	-.12	.05	0.9	-1	0.9	-1	.51	6	item6
1768	704	2.5	2.52	.92	.05	1.1	2	1.1	2	.29	7	item7
2435	697	3.5	3.56	-1.05	.06	1.0	0	0.9	-1	.42	8	item8
2368	705	3.4	3.43	-.70	.06	1.0	0	1.0	0	.33	9	item9
1434	676	2.1	2.09	1.58	.05	1.4	7	1.5	8	.13	10	item10
2158	675	3.2	3.26	-.33	.06	0.9	-2	0.9	-1	.33	11	item11
2087.5	694.4	3.0	3.05	.00	.05	1.0	-0.1	1.0	0.0	.35	Mean (Count: 11)	
272.0	11.2	0.4	0.40	.70	.00	0.2	3.3	0.2	3.5	.10	S.D.	

RMSE (Model) .05 Adj S.D. .70 Separation 12.93 Reliability .99
Fixed (all same) chi-square: 1862.4 d.f.: 10 significance: .00
Random (normal) chi-square: 10.0 d.f.: 9 significance: .35

The issue here:

- Assuming the two projects were independent: we have two valid (separate) measures of students' perceptions of their pre-university mathematical teaching ➔ no problem

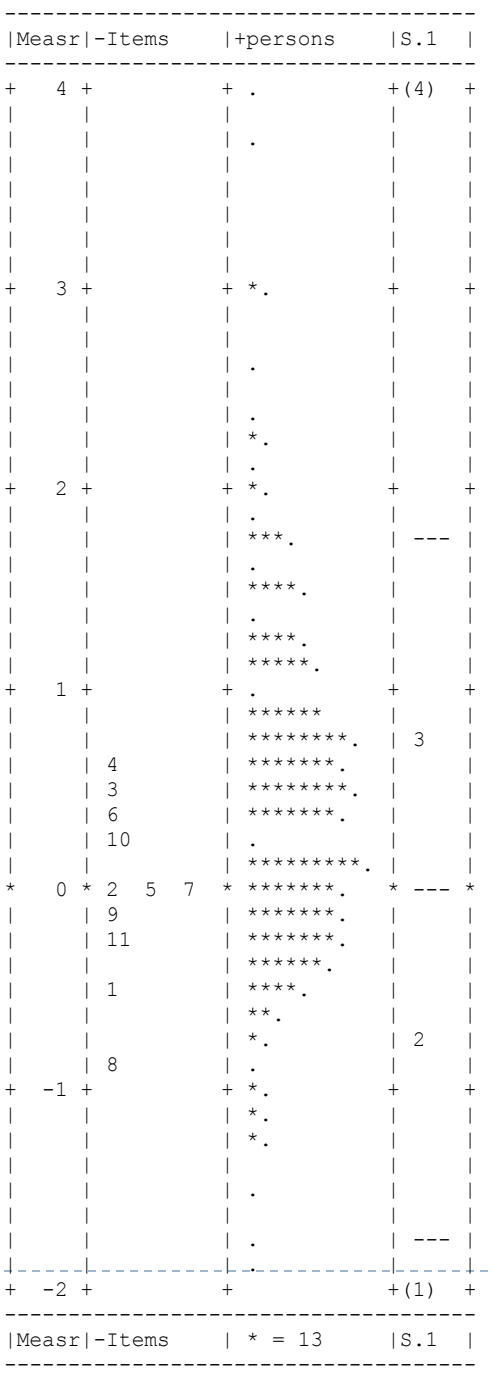
- BUT: if we were to link the data of the two projects and proceed with comparative statements → more needs to be done

- In Rasch (measurement) terms: we need to explore and deal with DIF

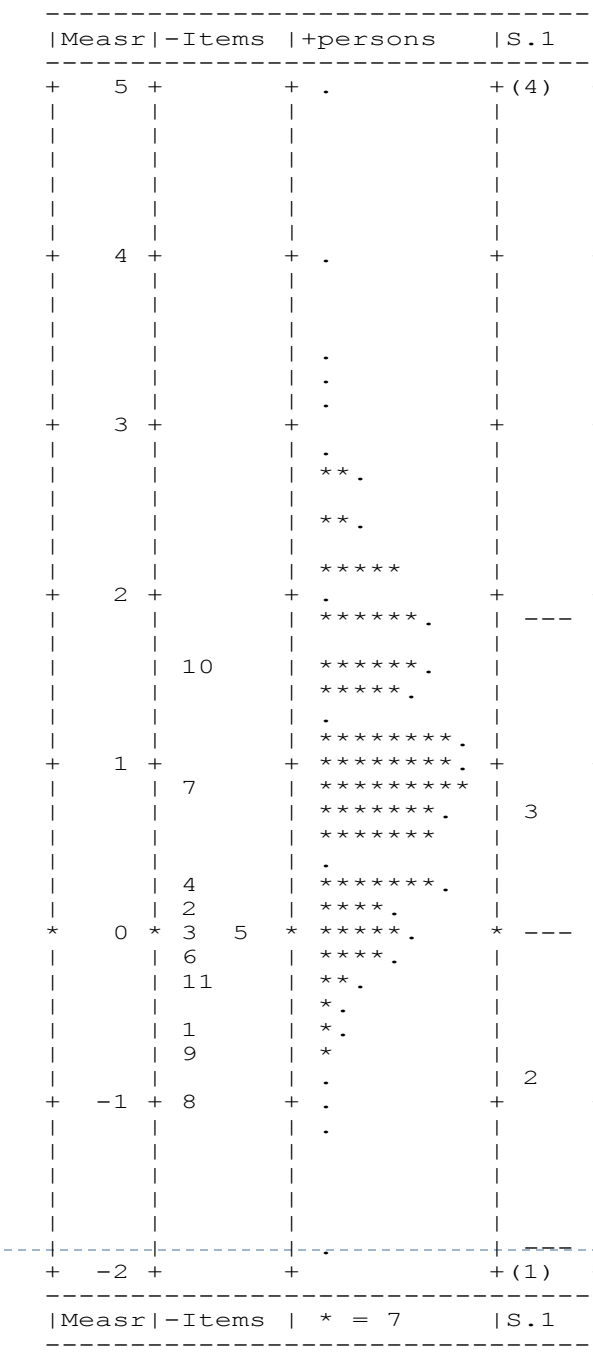
Differential Item Functioning (DIF)

- When a variable is used with different groups of persons [or to measure the same persons on different occasions], it is essential that the identity of the variable be maintained from group to group.
- Only if the item calibrations are invariant from group to group can meaningful comparisons of person measures be made.
- Differential Item Functioning (DIF): a statistical way to inform this process
- DIF measurement may be used to reduce this source of test invalidity and allows researchers to concentrate on the other explanations for group differences in test scores.
- Groups here: Students from UK and Norway

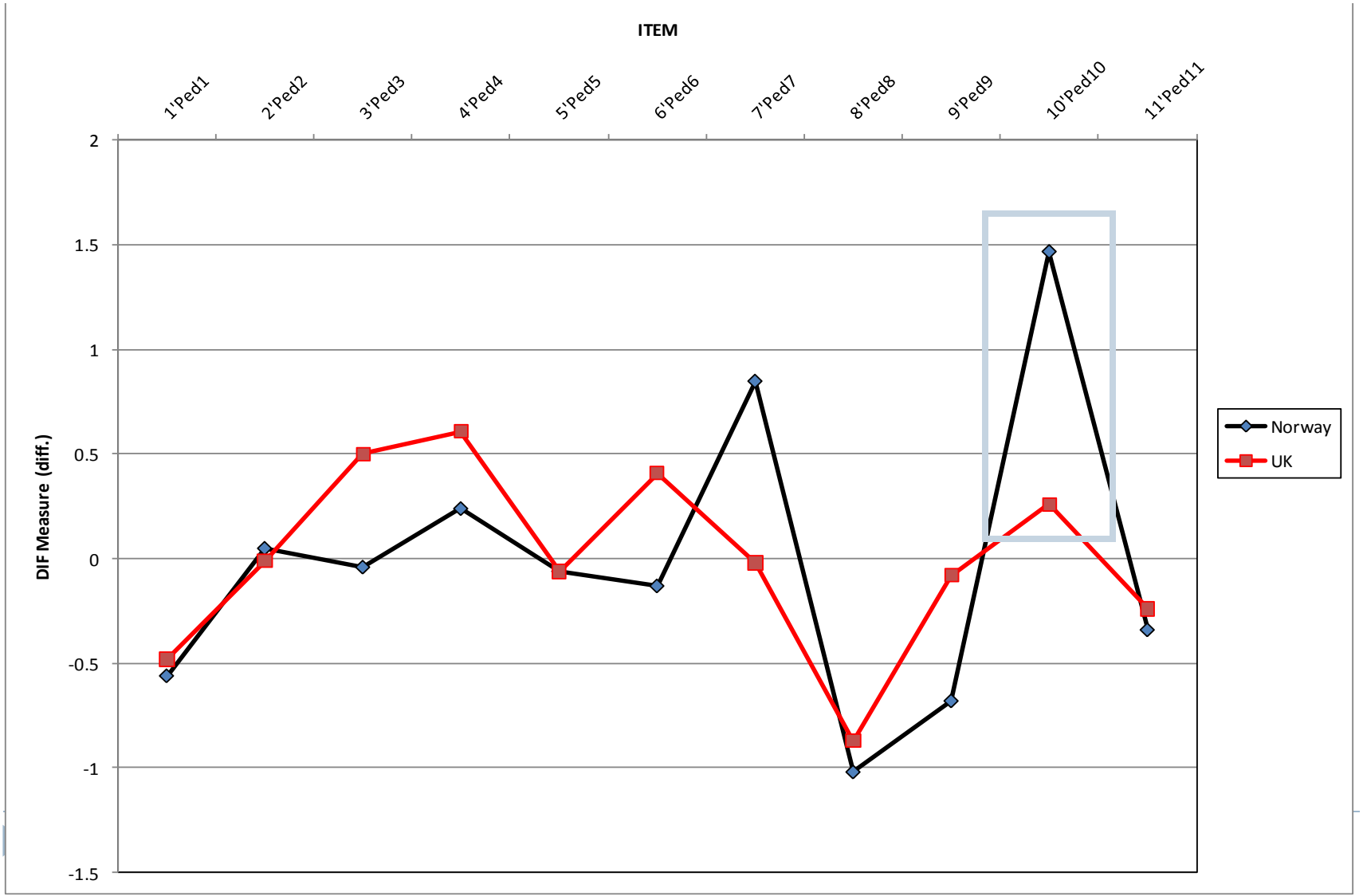
UK


|Measr|-Items |+persons |S.1 |
+ 4 +
+ 3 +
+ 2 +
+ 1 +
* 0 * 2 5 7 *
+ -1 +
+ -2 +
|Measr|-Items | * = 13 |S.1 |

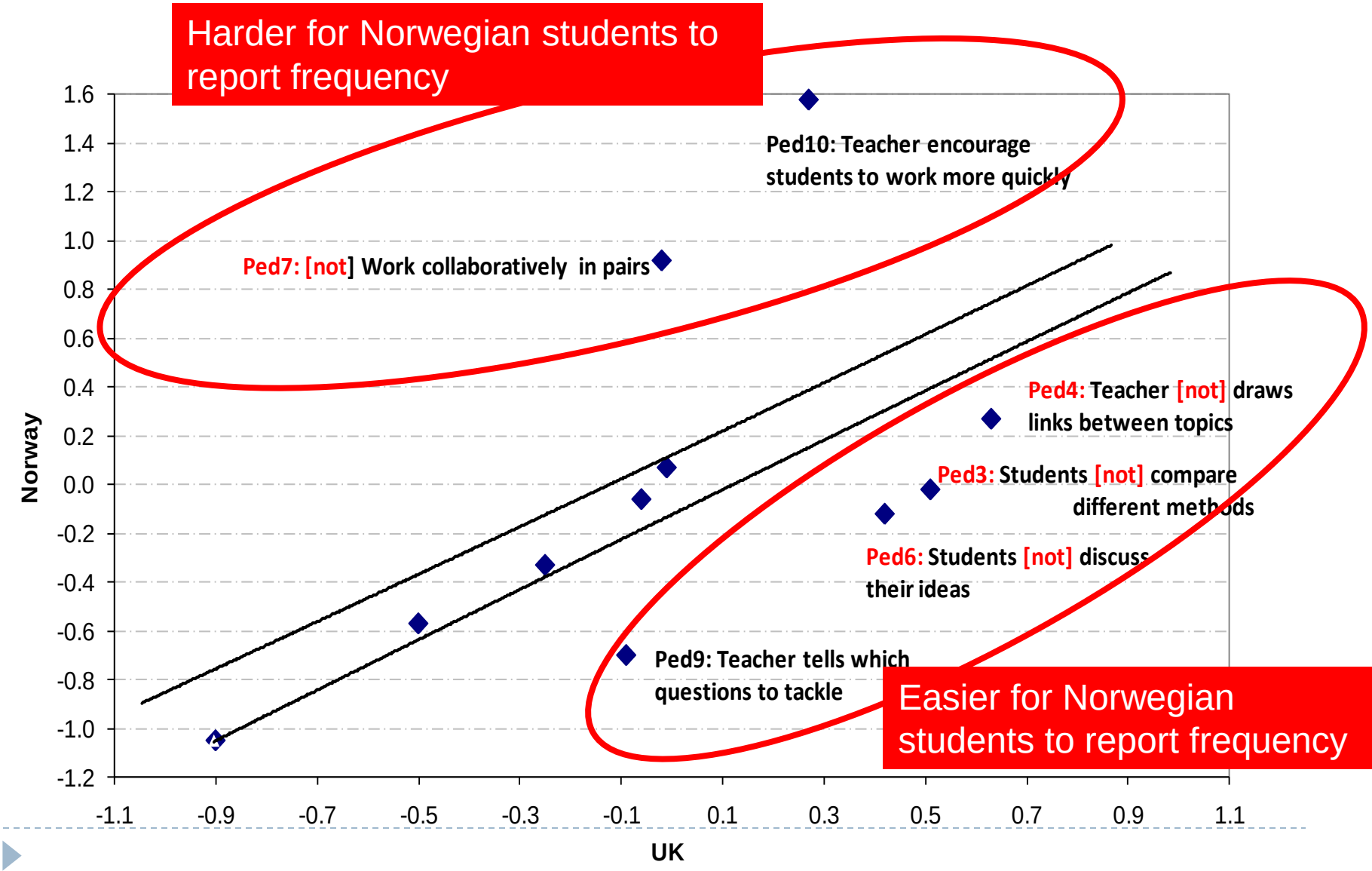
Norway


|Measr|-Items |+persons |S.1 |
+ 5 +
+ 4 +
+ 3 +
+ 2 +
+ 1 +
* 0 * 3 5 *
+ -1 +
+ -2 +
|Measr|-Items | * = 7 |S.1 |

Differential Item Functioning (DIF): Item measures of the two groups



Differential Item Functioning (DIF)

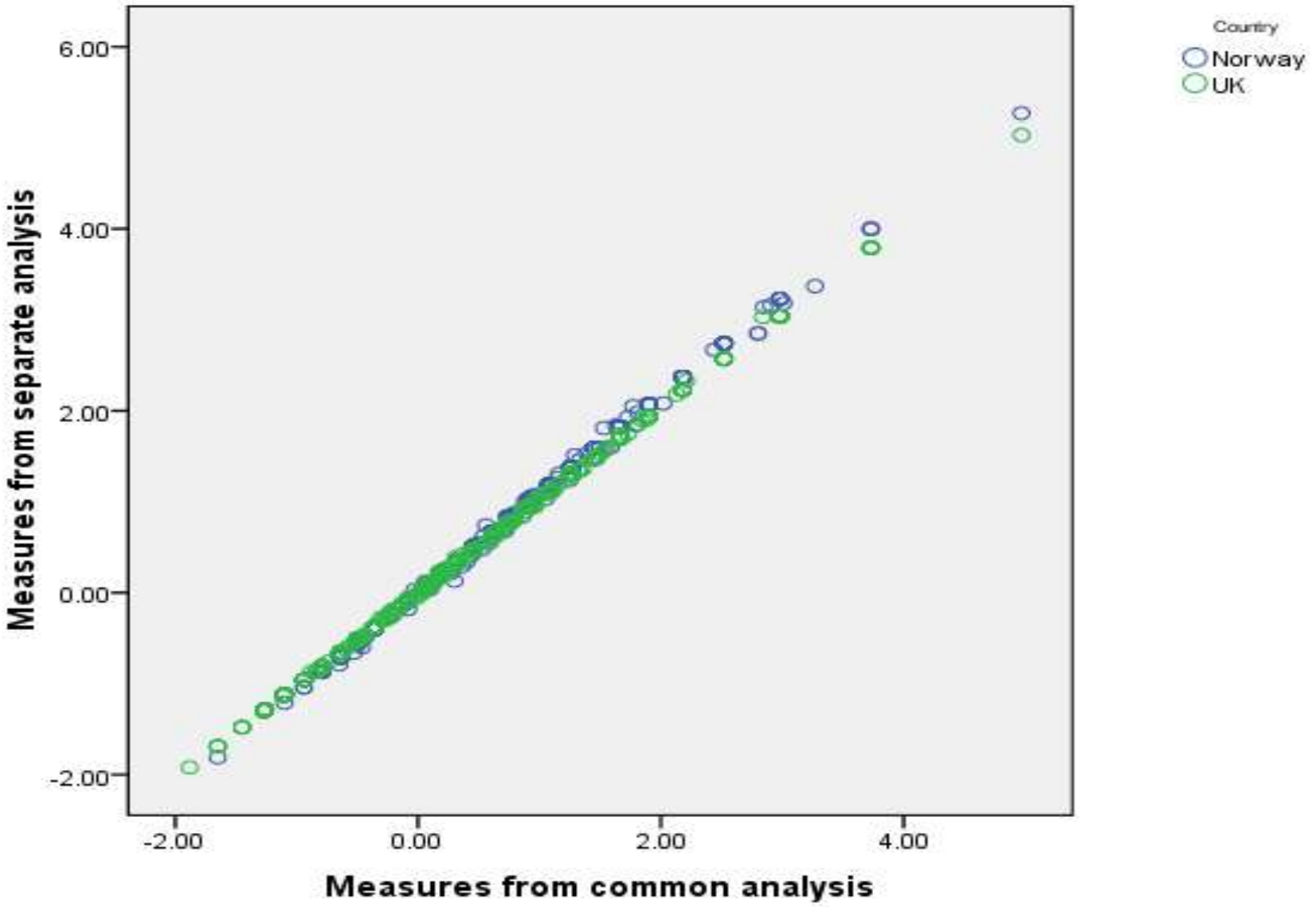


Differential Item Functioning

- DIF refers to a psychometric difference in how an item functions for two groups. DIF refers to a difference in item performance between two comparable groups of examinees, that is, groups that are matched with respect to the construct being measured by the test. The comparison of matched or comparable groups is critical because it is important to distinguish between differences in item functioning from differences between groups" (Dorans & Holland, 1993, p. 35).

- So the question remains: Is the instrument biased or differences are due to real differences?
- Or a more general question: When DIF becomes BIAS???

Plotting students' measures with two different analysis



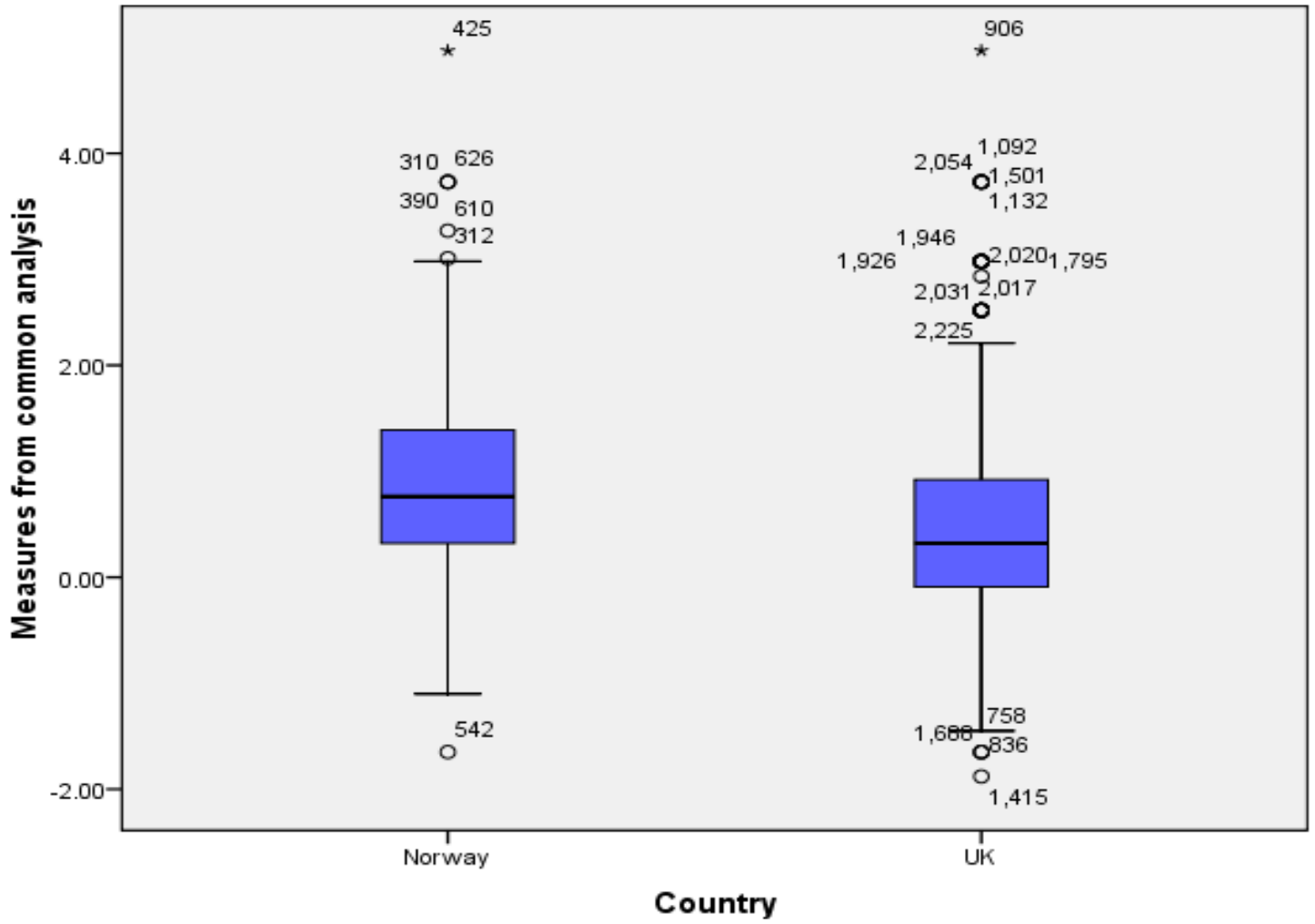
A comparative question

- Are the students from these two countries exposed to different pre-university practices?

(according to their report)

The Norwegian students reported more transmissionist practices in their pre-university maths courses

t=11.66, p<0.001



Another question

- How is this measure of students' perceived pre-university pedagogical experience associated with other measures of interest (e.g. dispositions, grades etc)

Some correlations from TransMaths UK

Pearson Correlations UK results	**Pedagogy at Uni**	**Pre-University Pedagogy**
Math Support at University (DP5)	Non significant	**-0.19 (p<0.05)**
Transitional Feelings (DP5)	**-0.20 (p<0.001)**	Non significant
Disposition to Finish Course_DP5	**-0.12 (p<0.05)**	Non significant
Math confidence (DP5)	Non significant	**-0.17 (p<0.001)**
MHE disposition (DP5)	Non significant	**-0.19 (p<0.001)**

The Teleprism student survey and some initial findings

Key	Never	Rarely	Sometimes	Always

Item name	Frequency bars
The teacher asks us questions.	
The teacher tells us which questions/activities to do.	
The teacher asks us to explain how we get our answers.	
We listen to the teacher talk about the topic.	
The teacher expects us to remember important ideas learnt in the past.	
We copy the teacher's notes from the board.	
The teacher gives us problems to investigate.	
The teacher asks us what we already know about a lesson topic.	
We discuss ideas with the whole classroom.	
The teacher uses the computer to teach some topics.	
We talk with other students about how to solve problems.	
We work through exercises from the textbook.	
We use calculators.	
We ask other students to explain their ideas.	
We explain our work to the whole class.	
The teacher tells us to work more quickly.	
The teacher tells us what value the lesson topic has for future use.	
We work together in groups on projects.	
What we learn is related with our out-of-school life.	
We learn that mathematics is about inventing rules.	
We get assignments to research topics on our own.	
The teacher starts new topics with problems about the world.	
We use computers.	
We do projects (assignments) that include other school subjects.	
We learn how mathematics has changed over time.	
We use other things like newspapers, magazines, or video.	

Some concluding points

- We showed how it is possible to measure ‘pedagogy’ across various stages of mathematics education
- (from Secondary School to University)
- Cross-national comparability

- Still to come:
- University pedagogy cross national comparisons (Norway - UK)
- Modeling of dispositions considering pedagogy (Norway)
- Student’s perceptions vs their teachers’ perceptions (Teleprism)
- Teleprism students’ perceptions from Year 7 to 11 with similar instruments…are they comparable or DIF causes problems??

References – for more information

- Pampaka, M., Williams, J. & Hutcheson, 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., Hutcheson, G. D., Wake, G., Black, L., Davis, P., & Hernandez-Martinez, P. (2011). The association between mathematics pedagogy and learners' dispositions for university study. British Educational Research Journal: First published on: 15 April 2011 (iFirst).
- Hutcheson, G. D., Pampaka, M., & Williams, J. (2011). Enrolment, achievement and retention on 'traditional' and 'Use of mathematics' pre-university courses. *Research in Mathematics Education, 13* (2), 147-168. D
- Williams, J., Black, L., Hernandez-Martinez, P., Davis, P., Pampaka, M. & Wake, G. (2009). *Repertoires of aspiration, narratives of identity, and cultural models of mathematics in practice*. In M. César & K. Kumpulainen (Eds), Social Interactions in Multicultural Settings (Chapter 2, pp. 39-69). Rotterdam: Sense.
- Williams, J., Black, L., Hernandez-Martinez, P., Davis, P. Pampaka, M. & Wake, G. (2009). *Pedagogies for social diversity and difference*. In M. David et. al. (Eds), Improving learning by Widening Participation in Higher Education (pp. 109 -123). London: Routledge.

- More information: maria.pampaka@manchester.ac.uk