Statistical versus personal narratives of the effect of pedagogy on learning

Pampaka, M., Williams, J. & Wake, G.

Abstract

In this paper we reflect on the 'effectiveness research' that the project has conducted as regards 'pedagogy' we reconceptualise this work as a form of 'narrative': in itself, and contrast it with the narratives of our qualitative research, and with a reflective story of the research written for the purposes of this paper. For us 'story-telling' is an activity of reporting that, according to activity theory, has an object, but is mediated by instruments (genres) and rules (story-telling norms etc) that serve certain, sometimes inexplicit, purposes in regards to the outcomes subjectively sought. In this way we try to generate insights into our research, and hence perhaps research in general.

Introduction

This paper is concerned with 'narrative' and 'pedagogy' as much as the other three in this symposium, but from a different point of view. We consider here the stories research tells of pedagogy as cultural narratives (Bruner, 1996) which are 'told' to an 'audience' from whom an outcome is desired (Engeström, 1995), using an appropriate narrative 'genre' (Bakhtin, 1986).

Also, like the others within this symposium, we draw on data from the ESRC-TLRP research project in widening-participation in HE: 'Keeping open the door to mathematically-demanding Further & Higher Education (F&HE) programmes' (see http://www.tlrp.org/ and http://www.lta.education.manchester.ac.uk/TLRP/). However, here we are interested in the research products as data, that is we reflect on the publications and papers of the project regarding pedagogy and analyse them as narrative constructions, where narrating a story is understood as an 'activity' that researchers may engage in individually or collectively as much as any other narrator.

We will consider our 'effectiveness' research in which we developed analyses of pedagogy at a macro, cross-institutional level, by (a) constructing a measure of pedagogy and (b) analysing its effect on the learning outcomes of c1700 students. The projects' research output from this analysis argued for the validity of the constructs (within a measurement methodological frame) and measured the relationship between

this measure of pedagogy and the disposition of the students to study mathematics further (while also exploring the effects of other variables). The result is a statistical 'story' about the 'effects' of 'teacher-centrism' for a policy audience.

But the project also produced (more qualitative) papers that tell of the effects of different pedagogies on classroom activity, and on learners' dispositions. Some of this is told in other papers in this symposium (e.g. Williams et al., Wake et al.) and elsewhere (Wake & Pampaka, 2008; Wake et al., 2007; Williams et al., in press). Some of these narratives seem to contradict the statistical story. We, therefore, tell a story of the 'effectiveness' research from the insider point of view. That will include the project team's reasoning behind the (need for the) development of the pedagogical instrument that led to the 'statistical' stories. It will be the story of the research team's multiple subjectivities during the pilot case studies, and how these motivated us to capture a dimension of pedagogy that might be treated as distinct from the curriculum effect in our original proposal.

So, the main body of the paper is based on the presentation of some pictures/figures and their accompanying story:

- A story of assessment validity for a pedagogy measure and the statistical/ modelling story of students' outcomes and the effect of pedagogy;
- The story of the case study work that produces an account of the influence of pedagogy on learning; and
- A 'political' story of our research teams' work.

Each 'story' is mainly divided into three parts: (a) a brief outline description of the research methodology and findings; (b) an analysis of the narrative using "AT/CHAT" triangle that defines and decomposes the story within a socio-cultural theoretical perspective. The distinct natures of the 'statistical' narratives and the 'case study' narrative will be discussed: how is narrative constrained by methodological choices, and how should we to narrate research on pedagogy?

STORY 1 - The effectiveness research: measurement of pedagogy and its relation to learning outcomes

Part A: Measuring pedagogy

The first part of this 'statitistical' story is based on a summary of an assessment paper (originally an AERA paper, now submitted to a journal). In Pampaka et al. (under review) we report the construction and validation of the instrument used to measure teacher-centered-ness or 'teacher-centrism' of pedagogy in our very particular, widening participation, project context. We show how its method of development involved interactive processes of qualitative interpretation and statistical modelling, and the role this validation process played in helping our research to make substantive claims about widening participation in mathematics. An earlier version of this work was also presented to AERA (Pampaka, Williams, Davis, & Wake, 2008).

How was it done? (Methodology etc)

The work of Swan[1] (2006), which is conceptually based on the research findings of Askew et al. (1997) and Ernest (1991), was particularly appropriate for our study in post-compulsory, "advanced" level mathematics college classrooms. The decision to use the items of Swan's 'practice scale' (Swan, 2006) was mainly a practical one: the instrument was designed for a national development project of GCSE[2] algebra using almost exactly the same population of Mathematics teachers as our study. Hence, the items were highly appropriate for our target teacher population.

A revised version of the instrument was created after some amendments given a pilot analysis of Swan's data, and a constructive discussion of its functionality and usefulness within a teacher conference. The final version included 28 items presented to the teachers as statements asking them to report the frequency with which certain activities take place in their classroom (using a 5-point Likert scale). An example item is the following: "Students work through exercises" (Almost never, Occasionally, About half the time, Most of the time, Almost always). The instrument was sent to the teachers of

[1] Swan (2006) adapted the three components of Ernest's work that can be used to characterize 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)

[2] GCSE is a major end qualification for the compulsory phase of education in the UK (approximately for age 16)

the (approximately 1800) students involved in our cross-college student questionnaire survey, at the second, 'end of year' student survey point. Teachers were asked to complete one survey for each of the mathematics classes they teach, so that students could be matched to the corresponding teacher's practice.

Data for this section come from 110 'cases' of pedagogy, by 95 individual teachers, from 31 further education colleges in the UK. This sample includes 78 cases of Traditional AS Mathematics (ASTrad) and 31 cases of AS Use of Mathematics (UoM) pedagogy. The qualitative data came from interviews with teachers in five of the institutions involved in our project.

The data (teachers' survey responses) were analysed using the one-parameter Rasch rating scale model (RSM, with the FACETS software). The Rasch model was 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 (Bond & Fox, 2001; Linacre, 2003; Wright, 1999). The model allows the item difficulty of each question or statement to be based on the way in which an appropriate group of subjects actually responded to that question in practice. The model establishes the relative difficulty of each item stem in recording the development of an attitude from the lowest to the highest levels the instrument is able to record (Andrich, 1999; Bond & Fox, 2001; Wright & Mok, 2000).

The cyclic process of analysis and interpretation of results involves accumulation of evidence to support validity arguments, through a validation process which is governed here by the guidelines summarised by Wolfe and Smith Jr, (2007a, 2007b) based on Messick's (1988, 1989) validity definitions. In particular we employ a measurement perspective on the issue of validity, drawing on the means provided by Rasch modelling, and interpreting our statistical results with qualitative evidence and judgement. To cut a long story short we now give the outcome of this process, i.e. the measure of pedagogy, and leave a (shortened) validity narrative for afterwards.

What is the aim, 'object' and outcome?

Figure 1 shows the resulting measurement scale (Picture 1). At the right side of the figure, the distribution of the teachers' pedagogy measures is shown as a histogram. The

higher the place of the "practice", the more teacher centered, or transmissionist, the pedagogy. Pedagogy that is mainly student-centered is at the bottom and pedagogy that is mainly 'teacher-centered' is at the top. On the left hand side of the figure the items that constitute the scale are presented, ranging from those easiest to report as frequent to the most "difficult" to report being frequent[3].

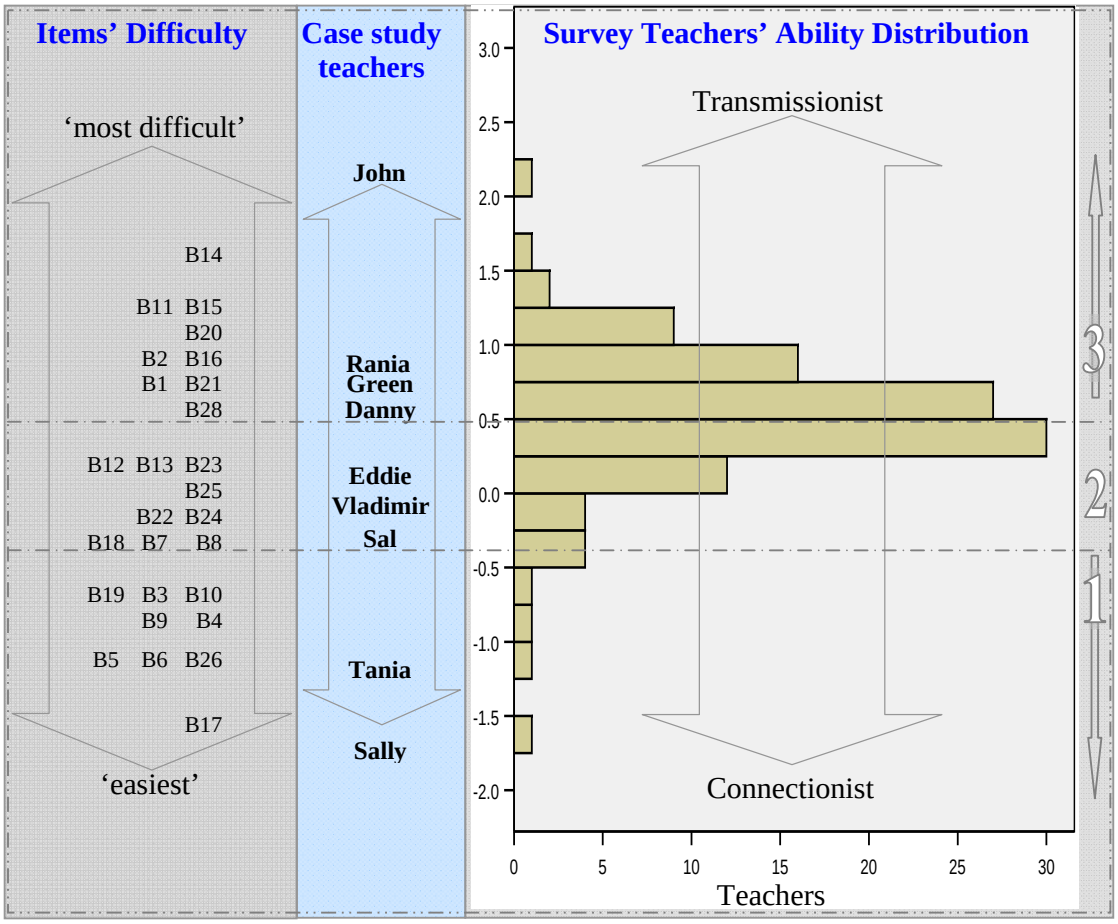


Figure 1: Picture 1- The 'transmissionist measurement scale'

Three hierarchical levels of this 'teacher centerness' measure can be distinguished (as shown on the right side of Figure 1, by the dotted lines) based on both the statistical results and the qualitative analysis of the homogeneity of the item content. The statistics that support this hierarchical progression of our measure, based on the Rasch modelling, are the person and item separation indices which provide estimations of the persons or

[3] The items at the end of the scale are presented as illustrative examples:
B14: I tend to follow the textbook closely
B17: Students [don't] invent their own methods
For this analysis and in order for results to be meaningful, the scoring of some items was reversed [e.g. B17]

items on the measured variable (Wright & Masters, 1982). A qualitative analysis of the items also resulted in just three categories giving three levels:

- Level 1: Frequently student-centered, more connectionist practice
- Level 2: Involves teacher practices from both ends of the spectrum in moderate frequencies
- Level 3: Teacher-centered, transmissionist, fast paced, exam orientated teaching

In the centre of Figure 1, the location of the pedagogies found in our case study colleges are tested to check for the validity and meaningfulness of our measures. Each pedagogy is identified with teachers we interviewed and observed teaching. It should be noted that the two teachers who define the ends of the scale (maximum and minimum score) are within our case study sample and their pedagogies and institutions have been analysed and reported (Williams et al., in press, in this symposium, also in Farsworth's ISCAR symposium).

Within the paper (Pampaka et al., under review) we detail the validation process by providing various type of evidence. Here we sum up the measurement story by listing the accumulated validity evidence to illustrate what is involved when saying such a story:

(i) *Item fit statistics:* In the Rasch context fit statistics indicate how accurately the data fit the model, and this, as a common practice, implies fulfilment of the unidimensionality assumption, hence suggesting development of one-dimensional scale. In our analysis, we came across some slightly misfiting items, which either we excluded or decided to leave within our measure based on theoretical reasons (also a common practice) or justification with qualitative data.

(ii) *Category Statistics* were also examined for the appropriateness of the Likert scale used and its interpretation by the respondents (Linacre, 2002), to justify what is usually called as communication validity (Lopez, 1995; 1996). These statistics indicated inappropriate use of a response category, and revision of the rating scale showed that a 4-point Likert scale will be more functional for future use of the instrument.

(iii) *Person – item maps and the item difficulty hierarchy* provided evidence for substantive, content and external validity (Figure 1).

(iv) *Differential Item Functioning (DIF) Analysis* was also essential for this study in order to establish the validity across the two different settings of pedagogy: UoM and ASTrad classes (Thissen, Steinberg, & Wainer, 1993; Wright & Masters, 1982). DIF was found for some items and (even though the issue was resolved by removing certain items) the final judgement made was that this might be due to 'real' differences in the teaching practices, afforded by the curriculum.

(v) Qualitative data from various teachers' interviews were also employed for triangulation. In Pampaka et al (under review) we give some extracts from the interviews with four of the teachers to illustrate how their narratives match the measures obtained by their self-reported surveys and elsewhere we provide a detailed contradiction of the narratives of the two extreme teachers (e.g. Williams et al, in this symposium).

Analysis within a CHAT framework

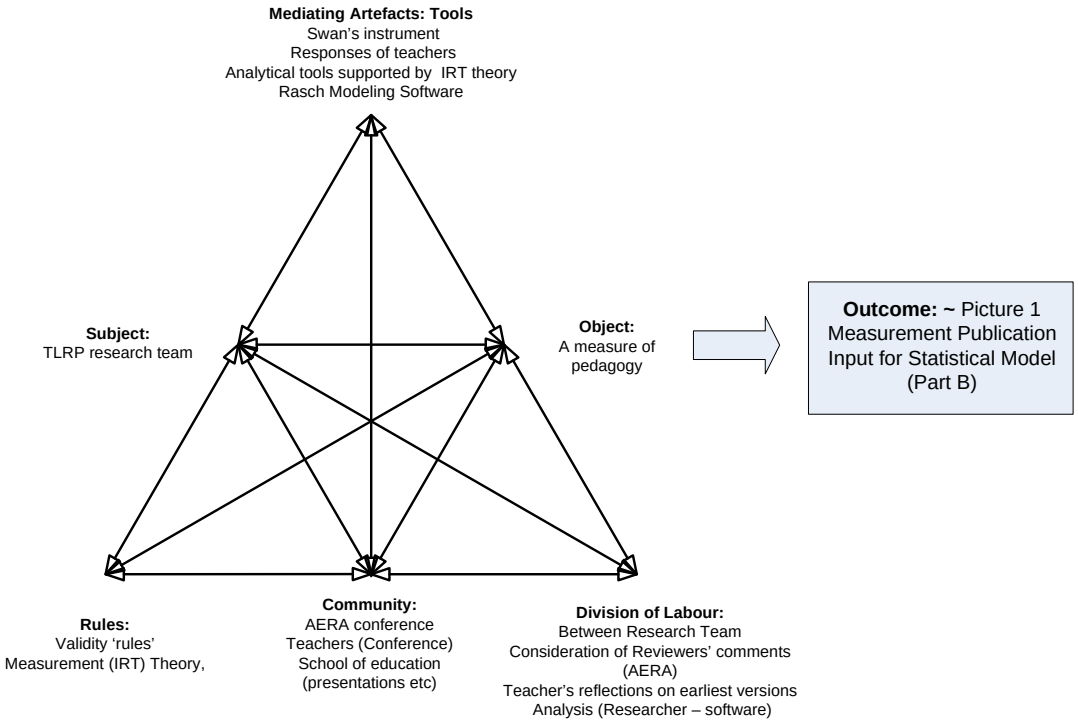


Figure 2: CHAT analysis of the measurement part of the story

What we should point out here is that even though teachers are 'voiced' and some had the chance to react and comment on the measure and its validity, the methodology implies that the teachers involved are objectified. So in a way, they are alienated from

the research and are reduced to a single number (i.e. their pedagogic measure) and positioned on the line (of Figure 1).

We also note that the 'narrative' of this research calls upon 'methods' of validation established within the measurement research community that calls for a certain quite specific 'standard of rigour'. It is this which allows for academic publication outcomes (a need of our project as a whole as well as the individuals within the team) but also which is a passport for the research to be taken seriously within an 'effectiveness' culture (to be examined later).

Part B - Statistical Modeling of learning outcomes

The previously described pedagogic instrument allowed us then to measure with some confidence the student- or teacher-centredness of pedagogic practices. The measure was then used to model the impact of pedagogy on learning outcomes within our project. Here we will present another statistical picture and the continuation/end of the previous story with (an example) model where the influence of pedagogy was found to be important on students' learning outcomes.

How? (Methods etc)

In order to fulfil the initial purpose of the constructed measure (i.e. to match pedagogy with the students' learning outcomes), a score of 'pedagogy' was given for each student based on their teachers' responses to the teacher survey (or average pedagogy score when a student has 2 or more teachers for mathematics). Data from students for analysis reported in this section come from two data points [beginning (N=1792) and end (N=1049) of their AS course/year]. The following results are based on the 750 students (475 AS Mathematics, 275 AS Use of Mathematics) for whom we have the matched learning outcome data and teachers' pedagogy measure.

Data analysis employed Generalised Linear Modelling of the student's learning outcome variables[4] over different time intervals (Hoffman, 2004; Hutchenson & Sofroniou, 1999) based on the following framework (Pampaka et al., 2007):

[4] The development and validation of these measures is reported elsewhere (Pampaka et al., 2007; Wake & Pampaka, 2007; Williams et al., 2007).

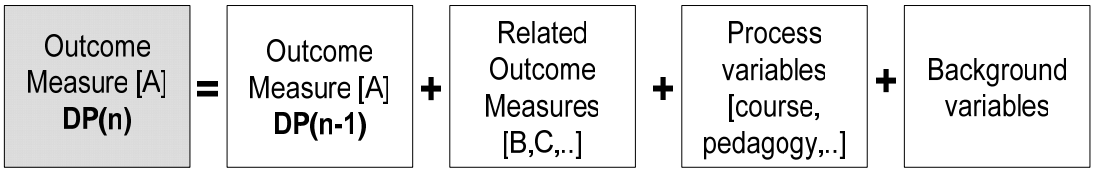


Figure 3: Process for modelling outcome variables at data point (n) regressed on data point (n - 1)

This framework allowed us to model the effects of background variables and earlier 'input' measures (including outcomes from *previous* time intervals) and conditions (e.g. 'Traditional' Vs 'Experimental' programmes and pedagogy) on the outcome measure (disposition or attainment).

What is the outcome?

Picture 2 (Figure 4) presents the resulting regression model for the value added to maths-disposition (i.e. disposition of the student to continue with mathematically demanding subjects at Data Point 2- DP2) and the effects plots of the explanatory variables on the response variable, for easier interpretation (Hutcheson, 2008).

The figure shows the statistically significant negative effects of pedagogy on students' dispositions to study mathematics further. The addition of background and other variables (i.e. GCSE grade and course) did not produce any other statistically significant effect, and also did not cause any change to the effect in the pedagogy measure. Overall we find that the pedagogy has a further negative effect, that is, transmissionist or 'teacher- and subject-centred teaching' is likely to further depress the students' maths-disposition.

A Regression Model for HE maths-disposition at DP2 (MdispDP2)

	Coefficient B	s.e.	t	p
(Constant)	-0.91847	0.08661	-10.605	< 2e-16
MHEdisp1	0.59801	0.03460	17.284	< 2e-16
MSE-2	0.34031	0.04706	7.232	1.25e-12
OtherSubjectsMathsDP2	0.14425	0.03073	4.695	3.21e-06
AveragePed	-0.24701	0.07966	-3.101	0.00201

F (4, 702) =127.6, p<0.001, R^2=0.421(Adjusted R^2=0.4177)

The effects plots of the explanatory variables on the response variable

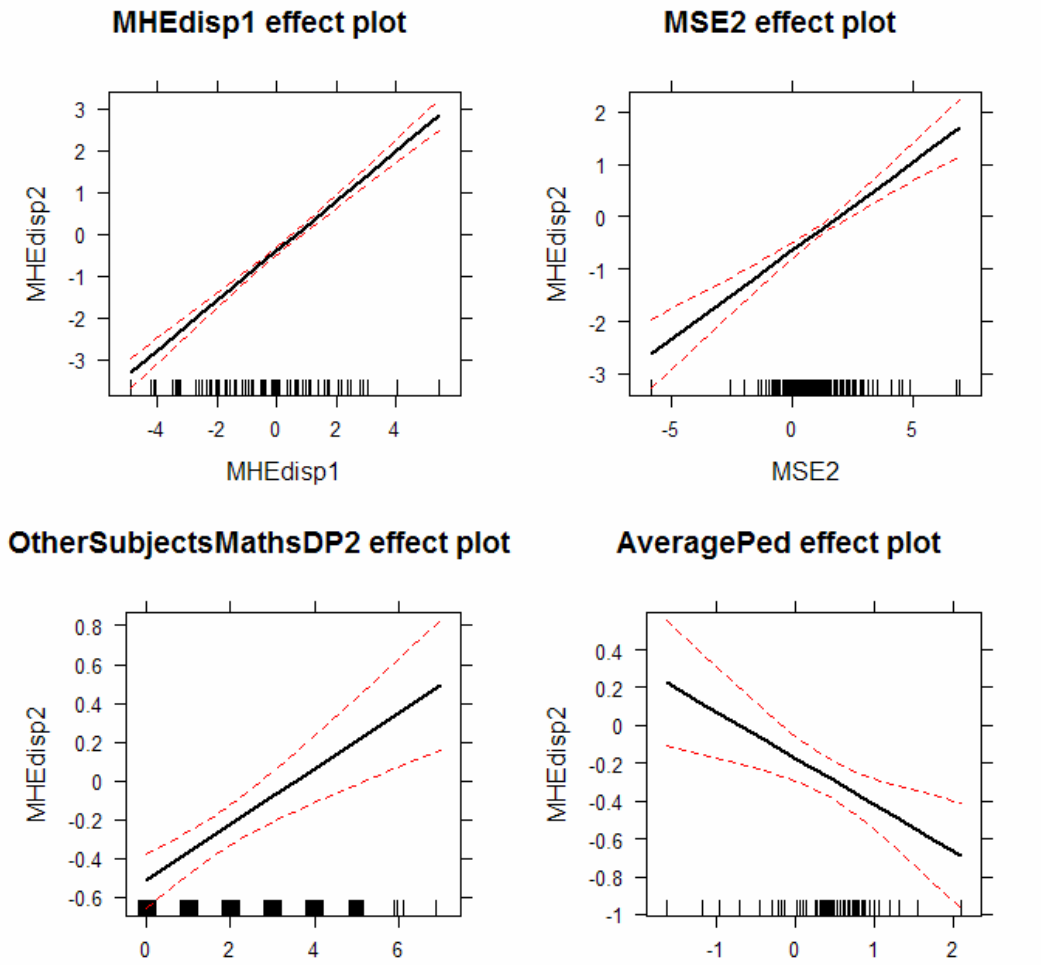


Figure 4: Picture 2- A model for maths disposition

In the paper (Pampaka et al., under review) we wondered about the educational significance of this slope/effect. This slope corresponds in effect to a difference of approximately the same magnitude as the difference between dispositions of students doing the two programmes (those following the traditional course are of course relatively highly disposed to do further mathematical study, as the UoM is a terminal course that does not allow students to progress to a full A-level, and these scores are

consistently higher before and at the end of the course). This should be claimed as educationally significant, if dispositions to study more mathematics are regarded as important learning outcomes.

This effect, however, disappears when the most highly connectionist few classes are removed from the data, implying that most of this effect was caused in these few classes (see next story). An important result is that there is no significant additional effect of pedagogy on grades, or dropout in any value added model.

Analysis within a CHAT framework

Returning to our analysis of this activity as a narrative genre, we note that the quantitative, statistical modelling offers the research credibility within the 'effectiveness' research community and the policy makers for whom such research is a minimal requirement, the acme of 'standards' in research (at least in government here and in the US) (e.g. BERA, 2004, http://www.bera.ac.uk/publications/guides.php, APA, AERA publication standards, http://www.apa.org/science/standards.html).

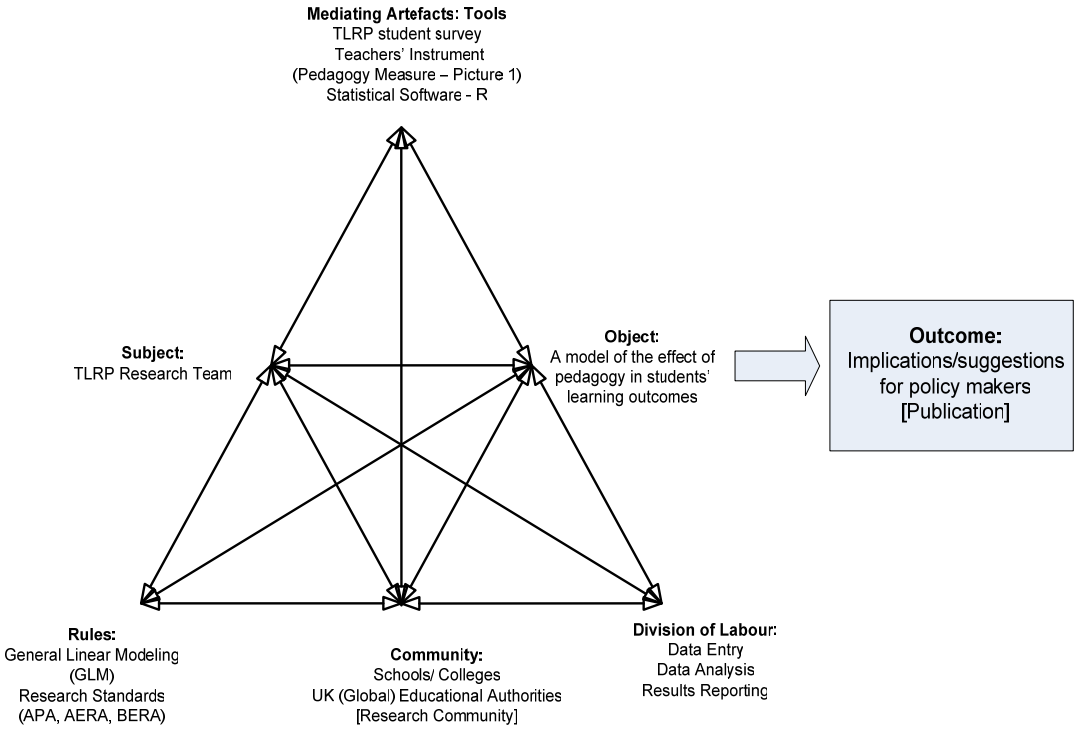
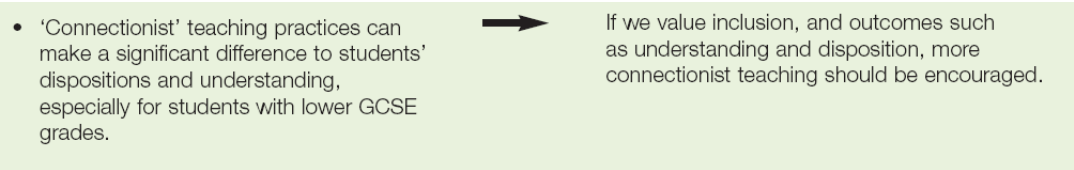


Figure 5: CHAT analysis of the modelling part of the story

The rules of the game, research standards in this case, include among others 'objectivity' of the research, the unanimity of the research team, validity and ethical considerations. Having followed the rules of the game, we are then 'allowed' in this community to make recommendations to policy makers.

A conclusion of our findings, therefore, stated that programme and pedagogy can make significant differences to learning outcomes for these students, especially in terms of drop-out and the disposition to continue to study mathematics, sometimes despite injurious policy and institutional influences. Here is how this was presented in the briefing document (Williams et al., 2008), as one of the three implications to policy makers:



And also a concluding quote:

> Teachers' practices (mostly transmissionist in this study) were powerfully shaped by the culture of performance, and by institutional factors such as the College access policy. Tests, exam grades and league table performance were dominant in shaping practice, though in particular circumstances we found some space for teachers' professional identity to provide agency so that some connectionist practices were able to survive. Reducing the pressure for immediate test performance and league tables could make room for teachers' professional development.
> (Williams et al., 2008)

Note that implications for policy here make many assumptions, like:

- Of course, that the correlation is causative;
- that it will be possible to manipulate the independent variable, so as to achieve the desired changes in the dependent variable.

These assumptions are examined to a certain extent in the qualitative research (most quantitative research of course does not conduct such coordinated research.) which follows.

STORY 2: Pedagogy within a qualitative analysis perspective

How? **[*College case studies and lesson analyses*]**

Our case study work in Colleges consisted of interviews with teachers and managers, videoed lesson observations, analysis of texts and artefacts, student interviews, and later re-visits to teachers to test findings. Our analyses of lesson observations included microanalyses of talk, analysis of 'whole lesson' narratives, and analyses of mathematical discourses that are best conceptualised on longer timescales (over the whole Programme, and the place of mathematics in the life course).

Each College was initially treated as a case study in which a combination of Programme, departmental culture/ approach, community/geo-demographics, institutional culture and market-position, provided a distinct context situating the phenomenon of teaching and learning mathematics. The interviews with teachers and College managers and analyses of the Mathematics Programmes provided us with insight into how the teaching approach came to be structured or mediated (object of activity, pace, interactional norms, tools and texts).

What is the outcome (some of the outcomes)?

Key differences between experiences in different contexts were identified and described: lessons were analysed as evidence of the social practice of teaching and learning, and cross-referred to independent interview accounts by teachers and learners. Differences due to pedagogy (transmissionist versus connectionist) were also found and presented (see http://www.lta.education.manchester.ac.uk/TLRP/).

Here we present two such accounts (from other publications) to illustrate how pedagogy was dealt with and narrated from a more qualitative perspective and essentially by another researcher as the main author/narrator. The first one is coming from a paper presented at BERA 2007 (Wake et al., 2007) and the second from a paper recently presented at PME (Wake & Pampaka, 2008).

The first narrative comes from lesson analysis, and as became 'common practice' in this paper we introduce it with a picture: in this case the two dimensional analytical framework of Picture 3 with one dimension taking into account the teacher's narrative about the mathematics itself; (ii) with a second dimension which is socially focused that

takes account of the pedagogic practices that the teacher uses and which reflect the culture of mathematics teaching and learning within which the teacher works, together with a social discourse.

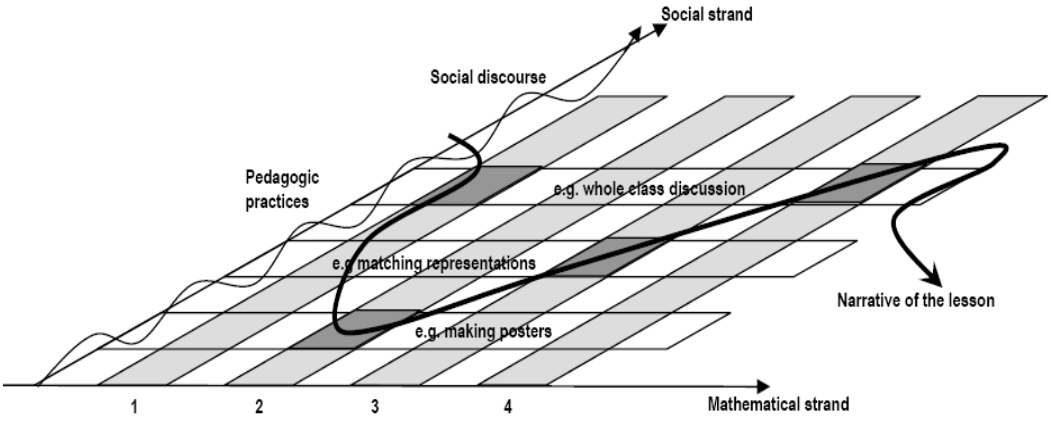


Figure 6 : Picture 3 – Framework for narrative analysis of lessons (Wake et al., 2007)

The following extract shows how pedagogy was described within this framework, and for this audience:

> The mathematical strand is driven by the mathematical argument that the teacher wants to present and reflects the way in which the teacher understands how mathematical ideas and processes familiar to his or her students may be (re-) introduced and interconnected to develop new (to the pupils) mathematics. On the other hand, the social strand contains references to 'why' as the teacher draws on a range of practices and discourse with which he or she attempts to motivate and engage his or her students in learning.
>
> Here, in two lessons, we exemplify how the operation of two different teachers in their classrooms can be analysed in this way. Both social and mathematical strands demonstrate the key features of narrative as the teacher engages the class in the development of a unique revelation of new mathematics. The 'social' narrative, in particular, we suggest is extremely important in engaging the learners not only through different activities but also a discourse that can motivate learning and which can also ensure, to a greater or lesser extent, connectivity with the mathematics itself. This appears potentially important as students attempt at the time, and later during periods of reflection, perhaps as they practise newly learnt techniques and so on, to make sense of the place of this new mathematics in the grand scheme of the discipline. Indeed, in Lesson 2 we observe that, much as in performance art, the audience (in this case the students), become part of the narrative themselves: the social strand of narrative ensures they are fully incorporated into the development of the mathematics itself. Importantly this is planned by the teacher from the outset, who although having key episodes in the mathematical development that she wishes to ensure are 'revealed' as part of the overall narrative, is willing to be flexible in this regard to ensure that the students themselves co-construct the narrative of the lesson. This contrasts, on the other hand, with Lesson 1, where the interaction of the students with the narrative is weak: although the teacher introduces a relatively prominent social strand, relating to his worms, he does not incorporate pedagogic practices that might ensure his

> students are engaged with this. Consequently, we argue, they may not be actively engaged with the mathematical strand to the narrative. (Wake et al., 2007)

At another level of analysis we were also able to identify a contradiction between lesson being student-centre in some activity and teacher-centered at other times "with these different phases having different objects and being differently mediated (Wake & Pampaka, 2008):

> When Sally engages students in social participation (eg group work, making posters etc) she is working with their ideas, (mathematical and intuitive) and ensures that their agency is given expression. Her own role is one of monitoring / assessing and so on. She facilitates students in their attempts to solve problems or create explanations 'in their own words', although they may be adopting some of her mathematics too to the extent that they understand it. The object of such activity is the problem the students work on and it is understood that they are to 'have a go' with their own mathematical tools/concepts.
>
> However, central to Sally's teaching are episodes when she takes control of the key elements of the emerging mathematical narrative, when her students' misconceptions are addressed. Sally interweaves this narrative with the students' own mathematical productions as they emerge from their 'sociable' activity: however, she subtly ensures that priority is given in such episodes to the 'correct', or more advanced mathematics that she wants them to understand, and of which she is pretty much the arbiter/judge. The object of this activity, then, is effectively to construct some sense of the 'more advanced mathematics' of which Sally is the key mediator.

CHAT analysis

Once again what we presented as papers elsewhere, here became data to initiate a comparative discussion and a preliminary CHAT analysis around pedagogy (which is by no means finished here, but just a taster). An observation on the CHAT triangle in Figure 6 does not indicate a lot of differences when compared to the one in Figure 5 for the statistical model.

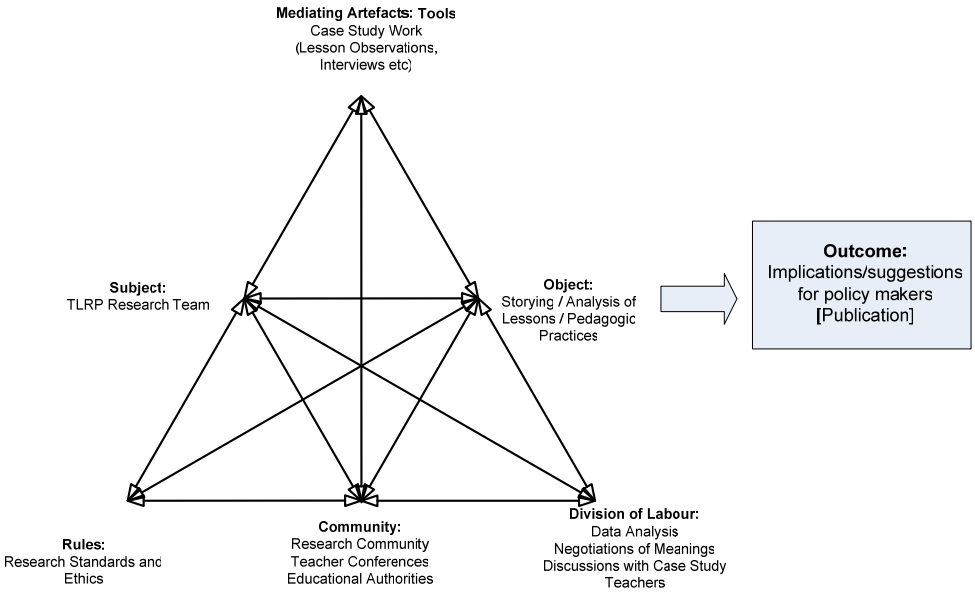


Figure 6: CHAT analysis of qualitative narratives of pedagogy

For both the statistical and the qualitative stories, (or the papers' re-conceptualisions if you prefer), narratives have to speak in a certain 'genre' to meet the requirements of being taken seriously by the intended communities. There is a common audience for both of these stories (i.e. the policy makers or the educational community) but there are also differences regarding the research 'audience' and the authors (maybe hidden).

We intend to consider this aspect in the complete analysis initiated in this working paper: we should consider the two (or more) communities here: the research community and the 'policy community', and the relations between these two. What implications does this have for 'sharing' methodologies and the positioning of teachers and learners as part of the object of research?

Off course this needs further discussion and a more detailed analysis however what would become and should be obvious is the difference in the two narratives, and the different agendas that people in the same research group may have, and how these fit together in a(ny) way. This issue is also captured in the next story.

STORY 3 - The story of the project team, its motives, object, and 'us'

The purpose of this narrative is to reflect on the project's work, its products and its collective 'identity' (in regards to the above-discussed products in particular)… The authors are not even sure if this account (first draft by Williams) is consensual. But the aim is to help clarify what may be 'hidden' in the aforementioned research accounts, and hence perhaps in any research account in the 'effectiveness' tradition. So here is again the picture: (some of) us at some point towards the end of the pilot phase (June 2006).

Pilot Case Studies:

TLRP Proposal:

"…it is suggested that appropriately designed programmes, perhaps UoM, may mediate pedagogies that can in turn influence learners' mathematical identities (especially for our target low-participation groups of students) differently than a traditional, usually didactic AS /A2 mathematics programme. However, we are aware that the two programmes (UoM and traditional) will not necessarily be implemented in distinctly 'didactic' versus 'inquiry-based, sociable' ways, and that the classroom culture is influenced by a complex of factors, including for instance teachers' beliefs, training and dispositions. Hence, although we expect to find distinct features of pedagogic culture associated with the programmes, we also expect variation in learning contexts and recognise the need to study the full complexity of a diversity of classroom cultures…"

Figure 7: Just Picture 4

In our (pilot) case study colleges we had the opportunity to observe teaching and learning in some considerable detail and use an analytic framework that includes dimensions of pedagogic practices and mathematical narrative to attempt to make sense of the ways in which teachers engage students in learning new mathematics. This allowed us to identify ways in which the different mathematics programmes might constrain or afford different pedagogies. In fact, we observed in pilot studies that while much of the traditional teaching we saw is 'transmissionist' and 'teacher-centred', there were teachers of traditional mathematics courses who described their teaching as 'connectionist' and whose practices were observed to be consistent with this description (Swan, 2006; Askew et al., 1997): the approach can be student-centred and involves students in conceptually-focussed discussion that is certainly 'communicative', but not necessarily use- or modelling- orientated. (See the descriptions of Wake et al., this symposium).

This observation posed a threat to the aims of the project, at least as perceived by some in the team. There was a hypothesis that the 'unconventional' UoM programme would be the one that might offer pedagogies that could make a significant difference to learning outcomes for less well-prepared students and so contribute to widening participation. Indeed we were quite sure from what we knew of the pass statistics on the two programmes that UoM allowed many less-well prepared students to achieve a successful outcome that would not (based on their prior attainment/grades) have been achieved on the traditional programmes in mathematics.

The threat however was that the difference between programmes – if we did into take account of the different pedagogies of the teachers – might not be interpretable in any meaningful way in terms of policy.

Yet here we were presented with the possibility that a significant number of classrooms were offering a pedagogy that might widen access/participation (as it is designed to do: see Williams et al., this symposium). The quantitative analysis planned then, could end up providing relationships that were disadvantageous to the 'alternative' curriculum, if this new aspect of pedagogy was not included in the modelling.

We therefore set about measuring pedagogic practice that might gain purchase on this dimension of widening participation: the result could then be a 'curriculum' dimension

(UoM versus traditional) as well as a ‘pedagogy dimension’ (traditional/transmissionist) versus alternative. We did not expect that these dimensions would be independent (we thought probably the UoM teachers would be less traditional than the ‘traditional’ programme teachers) but we thought that we would have a better understanding of any relationships involved if we were able to account for the two dimensions separately. We might then be able to make inferences such as ‘the effect of the pedagogy is …., but in addition to this, the re is a programme effect…’ or vice versa.

An even greater threat in prospect was that the two effects might, to a degree cancel out, and so by not measuring pedagogy we would produce results that suggested that the UoM programme was ineffective. This result would have been a disaster for at least two, perhaps three, of the project team (while others were less ‘interested’) who have invested a large part of their careers in developing ‘uses of mathematics’ as a means to improve mathematics education. Thus, in this story, the different subjectivities of the project team begin to play a role, and the ‘collective subject’ of this research begins to be called into question.

Indeed the ‘effectiveness’ research did produce some challenging findings, in that it revealed that there were few differences between the pedagogies of the teachers of the two programmes, but that there were significant differences between a few of the teachers on the traditional programme and all the others, and it was these few teachers that generated the (few, and very small) statistically significant results on dispositions.

Finally then the sense behind these results begins to emerge: the ‘effectiveness’ model that requires learning outcomes to be the product of Programme and pedagogy may be questioned. What if teachers’ habitus and their classroom pedagogy is in turn conditioned by the programme, the students’ prepared-ness, and the institution-community?

Conclusion and discussion

The ‘effectiveness’ research (story 1) stands on grounds of objectivity that make many assumptions about practice that are difficult to sustain. These can, to some extent, be ameliorated by qualitative research (story 2) – but such mixed research method is rare, and few research teams have the resources, skills and disposition to conduct such research.

Even then, the reflexive narration of the research (hinted at in story 3) may be necessary to understand the research process, and hence the outcomes reported in the previous two versions.

It is to be noted that any such reflexivity stands as a challenge to the ‘effectiveness’ research community, whose method involves a denial of any such subjectivity in research. The agenda of ‘what works’ that policy makers endorse requires at least (i) denial of subjectivity, (ii) downplaying of variability between pedagogues/learners, (iii) denial of ideology/politics.

The last of these is particularly interesting: the denial of ideology involves a promotion of common sense and the ideology and politics it hides (especially in policy-making!).

References

Andrich, D. (1999). Rating Scale Model. In G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 110 - 121). Oxford: Pergamon.

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

Bakhtin, M. M. (1986). *Speech Genres and Other Late Essays*. Austin, Texas: University of Texas Press.

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

Bruner, J. (1996). *The culture of education*. Cambridge, Mass.: Harvard University Press.

Engeström, R. (1995). Voice as communicative action. *Mind, Culture, and Activity, 2*, 192-215.

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

Hoffman, J. P. (2004). *Generalized Linear Models: An Applied Approach*: Pearson Education.

Hutchenson, G., & Sofroniou, N. (1999). *The Multivariate Social Scientist. Introductory Statistics Using Generalized Linear Models*. London: Sage.

Hutcheson, G. (2008). Dispositions towards studying Mathematically-Demanding subjects in HE. *TLRP-WP-Maths working paper*.

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

Linacre, J. M. (2003). *A user's guide to FACETS: Rasch-Model Computer programs [software manual]*. Chicago: Winsteps.com.

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.

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.

Pampaka, M., Williams, J., 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., Hutcheson, G., Black, L., Davis, P., Hernandez-Martinez, P., et al. (2007). *Measuring the 'effectiveness' of Programme and pedagogy on maths disposition and self efficacy measures*. Paper presented at the Paper presentation at the Annual Conference of the British Educational Research Association (BERA 2007).

Pampaka, M., Williams, J., Hutcheson, G., Davis, P., & Wake, G. (under review). Measuring pedagogic practice for widening participation in mathematics. *Assessment in Education: Principles, Policy & Practice*.

Swan, M. (2006). Designing and using research instruments to describe teh beliefs and practices of mathematics teachers. *Research in Education*(75), 58-70.

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.

Wake, G., & Pampaka, M. (2007). *Measuring Perceieved Self-Efficacy in Applying Mathematics.* Paper presented at the Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education (CERME) (WK 13, pp. 2210-2219), Larnaca, Cyprus.

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.*

Wake, G., Williams, J., Black, L., Davis, P., Hernandez-Martinez, P., & Pampaka, M. (2007). *Pedagogic practices and interweaving narratives in AS Mathematics classrooms.* Paper presented at the Annual Conference of the British Educational Research Association (BERA 2007).

Williams, J., Pampaka, M., Black, L., Davis, P., Hernandez-Martinez, P., & Wake, G. (2007). *Development and validation of two 'soft' outcome measures: Disposition to enter HE and disposition to study mathematically demanding subjects in HE.* Paper presented at the Paper presentation at the Annual Conference of the British Educational Research Association (BERA 2007)

Williams, J. S., Black, L., Davis, P., Hernandez-Martinez, P., Hutcheson, G., Nicholson, S., et al. (2008). *TLRP Research Briefing No 38: Keeping open the door to mathematically demanding programmes in Further and Higher Education*. School of Education, University of Manchester.

Williams, J. S., Black, L., Hernandez-Martinez, P., Davis, P., Pampaka, M., & Wake, G. (in press). Repertoires of aspiration, narratives of identity, and cultural models of mathematics in practice. In M. César & K. Kumpulainen (Eds.), *Social Interactions in Multicultural Settings* Rotterdam: Sense. Publishers.

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. (1999). Rasch Measurement Models. In G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 268-281). Oxford: Pergamon.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

Wright, B. D., & Mok, M. (2000). Rasch Models Overview. *Journal of Applied Measurement, 1*(1), 83-106.