

Raters' Understanding of Rating Scales as Abstracted Concept and as Instruments for Decision Making¹

Constant Leung and Alex. Teasdale
Centre for Applied Linguistic Research
Thames Valley University

Abstract

This paper reports on basic research into teachers' understanding and use of rating scales in the context of multilingual mainstream classrooms in the primary sector in England. The research employs a multi-stage interview procedure to gather data on teachers' use of assessment descriptors.

The findings indicate that teachers-as-raters draw upon a range of professional experience, personal interpretations and folk theories in arriving at judgements about children's language use. In informal on-going assessment of classroom spoken English there would appear to be no direct correspondence between pupil performance, the content of individual descriptors and teachers' justifications of allocation to grade. This raises fundamental questions about both the implementation and the outcomes of teacher assessment

1. Teacher Assessment of Speaking and Listening in the National Curriculum

The statutory assessment of speaking and listening in English of 7-year-old pupils at the end of Key Stage 1 (second year of elementary schooling) is classroom-based. Teachers are expected to assess pupils using the level descriptions published by the School Curriculum and Assessment Authority (1995). (See Appendix 1.) The assessment is based on teacher observation and evaluation of pupil language use in ordinary everyday classroom activities. In this study, we will be concerned with teachers' representations of assessment constructs

¹ Acknowledgements: This research was conducted with the assistance of a grant from the Centre for Applied Linguistic Research at Thames Valley University. The authors would like to thank Charmian Kenner for her help with the research interviews and the participating schools and teachers who gave generously of their time.

taking into account the multilingual pupil population of many of the English elementary classrooms.

2. Construct and Teacher Assessment

Within the psychometric paradigm, construct is often characterised in terms of the skill/s or attribute/s being measured (Gipps, 1994). The operationalisation of these skills is normally presumed to take place in a more or less stable environment whereby there is a high level of compatibility between the construct and the assessment instrument/test format. Where performance is judged, a high degree of compatibility is required between the actual construct and the rater's representation of it. Certain learning/assessment situations dispose themselves more favourably towards the harmonisation of components, eg the testing of grammar and certain approaches to the testing of reading. However, for Teacher Assessment in the National Curriculum, any such consistency or harmony between the different components of the process may be missing, or indeed not achievable because:

Teacher Assessment of speaking and listening is not operationalised in any specific format or condition; it is meant to be part of the ordinary teaching-learning process in the classroom. The NC level descriptors for speaking and listening are not relateable to any explicit account of construct which leaves teachers more than usually charged with interpreting the Descriptors. Teachers are advised that 'Teacher Assessment ... is based on a range of experiences and contexts and is intended to build up, over time, a full picture of a pupil's skills and understanding. By using a "best fit" judgement, teachers can recognise positive achievement and balance strengths and weaknesses in ways which enable these pupils' attainments to be recognised fully' (SCAA 1996: 18)

Whilst assessment is characterised as 'largely diagnostic, to inform plans for teaching', (SCAA 1996: 19), the fact that all second language pupils are put into age-appropriate teaching year groups [i.e. grades] to follow the mainstream curriculum together with English as a Mother Tongue pupils, irrespective of language background, current level of English language development and previous schooling experience, means that the assessment system must cope with diagnostics for both groups. From the point of view of

language assessment, there is a real potential of construct underrepresentation (Messick 1988).

3. Reasons for studying teachers' representations of constructs in the English National Curriculum

Given the lack of systematic standardisation in Teacher Assessment in the National Curriculum, there is a need to find out how teachers operationalise the scale and the system in general. Teacher Assessment is officially seen to serve two purposes: to diagnose learner needs (and to plan teaching) and to recognise pupil achievement. While the requirement for assessment for more than one purpose is by no means unique (see Gurn 1995 for a discussion), there is a need to find out to what extent these two purposes are reflected in teachers' practice.

More specifically, we need to understand the ways in which teachers identify as relevant aspects of ordinary curriculum tasks and other in-school activities involving language use since they may use such observations in formative assessment which impacts on their decisions about teaching. Furthermore, there is a need to find out the extent to which teachers' representations converge or diverge from the National Curriculum level descriptions. Teacher Assessment is a curious hybrid phenomenon: in the National Curriculum formulation, it places no requirement on the teacher/assessor to control task or performance conditions but relies in its outcomes on the consistency and validity of teacher interpretation of the descriptors (a process of linguistically encoding or recoding teacher representations). In the longer run, a knowledge of teachers' representations will help make explicit what teachers' criteria are and help form a critique of these 'folk' criteria. This is not only useful for teacher education and for specifying the professional knowledge base but also as an agenda for action in assessment, particularly in relation to provision for bilingual children.

Another reason for studying teacher's representations of assessment constructs is related to the issue of fairness. The NC level descriptions, it may be reasonably argued, represent an important component of the body of the subject knowledge as defined by the curriculum authorities. However, given the potential for variability in the assessment arrangements outlined above, we could

argue that it is necessary to investigate teachers' representations in assessment whether these are related to interpretation and use of the National Curriculum Level Descriptions or to other representations which are central to their judgements. In this way, we may be able not only to understand better what is happening in assessment but also to gain some insight into how the curriculum is realised. The latter point has particular salience because of the need for information for curriculum evaluation, the need to see how the curriculum at the classroom level is operated for the linguistically and ethnically diverse pupil population and the need to resource both teaching and assessment in ways that would ensure equality of entitlement and fairness of outcomes.

4. Espoused Theory and Theory-in-use

Initially, in conceptualising the research, two terms which have been used in studies of organisational behaviour were useful in clarifying our thinking. Schön (1987) revisits the terms Espoused Theory and Theory-in-use which originated in his earlier collaboration with Argyris (Argyris and Schön 1974). The account of these terms and the subsequent models built around them make assumptions about roles and behaviour in organisations which are not relevant to our study. However, metaphorically, the two terms were influential in shaping the research design. Schön (1987: 255-256) talks about the two terms in this way:

There are espoused theories which we use to explain or justify our behaviour... But there are also theories in use implicit in our patterns of spontaneous behaviour with others. Like other kinds of knowing-in-action, they are usually tacit. Often we are unable to describe them, and we are surprised to discover, when we do construct them by reflecting on the directly observable data, of our actual interpersonal practice, they are incongruent with the theories of action we espouse.

For our study, direct observation of inter-personal (or more critically, in our case, cognitive) behaviour in context would be unlikely to yield usable evidence. Our experience of using a repertory grid technique in stimulated recall interviews (Leung and Teasdale, 1996) and some of the work from the field of

Phenomenography² (Marton 1981; Säljö 1988) led us to consider the possibility of examining teachers' conceptions of assessment criteria in terms of abstracted non-applied conceptual networks (Espoused Theory) and applied concept in the situated action of assessment (Theory-in-use). Given the difficulties of data gathering, a mixture of the following elicitation formats was used: semi-structured interview; semi-structured interview with stimulated recall; and teacher audio diaries on their assessments.

As the study progressed, it became increasingly clear that the Espoused Theory and Theory-in-use distinction was not sustainable since the tendency towards narrative and exemplification in interview types blurred the distinction. Similarly, Phenomenography, which at the start of the research stimulated much intellectual questioning, became gradually less relevant to the conduct of the study as methodological issues, an area not well-defined in Phenomenography, became increasingly important.

The position we take is broadly consistent with Schutz (1970) and Berger and Luckmann's (1967) phenomenological views. Issues related to the nature and status of spoken discourse as tokens of mental or social representation are, of course, of key importance. Säljö's position is more or less consistent with our own. Intersubjectivity in the making and interpretations of meaning in social space would make a basic content analysis approach too simplistic and potentially too distorting of our data. In the following sections we return to some of these issues in addressing issues thrown up by the data.

² Phenomenography has its genesis in studies of learning conducted at the University of Gothenburg in the mid 1970s. It focusses on the nature of conceptualisations of specific phenomena. Much of the work has focussed on the nature of school and university students' understanding and learning within specific disciplines. Säljö (1988:42) in an important paper on Phenomenography comments that:

'... conceptions of reality are not considered as residing within individuals. In other words, people do not have specific conceptions of phenomena in the world around them in the sense that behavioural scientists have had a tendency to ascribe intellectual capacities or developmental stages to individuals. People may—and do—have a tendency to use particular conceptions of reality in a number of settings or in relation to a number of problems, but they cannot always be assumed to adopt that particular perspective on reality.'

5. The Current Research Study: specific research questions and overview of the design

In order to arrive at an understanding of teachers' professional representations of spoken language assessment and their application in the classroom, the following research questions were addressed:

- why and how do teachers assess spoken English?
- which descriptors in the first four level descriptions of the National Curriculum are regarded by teachers as important for assessing 7-year-olds and how do teachers interpret these descriptors?
- what National Curriculum level descriptors and other criteria do teachers use when assessing pupils?
- how and on what evidential bases do teachers arrive at their judgements?

The main subjects in the study were four Primary school teachers. Sixteen of their pupils, aged between 6 and 8, were also implicated in the study. These pupils appeared on edited video footage which was used in stimulated recall interviews with the teachers.

There were two stages of data gathering. In the first phase, video footage was shot in four classes³ in two primary schools using two cameras. The video footage consisted of naturally occurring classroom activities over 7 school days. In total, there were approximately 35 hours of footage. This footage was edited down to 16 3-5 minute segments of individual pupils engaging in a variety of learning activities. These individual segments were used later in interviews with the teachers. In the second phase, each of the four class teachers was interviewed three times over a five-week period. These interviews were semi-structured and each lasted between 60 -

³ Three of the four classes were in the same school; each of these Year 2 classes had their own class teacher and their own class space but they shared a large common central activity area. The fourth was a Year 3 class in another school. Given that data collection occurred early in the school year, it was felt that the Year 3 children constituted a representative sample for the purposes of the research.

80 minutes. The first interview focussed on teachers' reasons for and practice in the carrying out of Teacher Assessment of oral English. The teachers were also asked to identify what they held to be important descriptors within each of the National Curriculum level descriptions. No specific reference was made to individual pupils. In the second interview the focus was on specific pupils in the teacher's class (but see footnote 2 below). The main aim was to collect information on what teachers perceived as significant aspects of pupil language use and how such perceptions were represented in assessment terms. In the third interview, the teachers were shown individual pupil video segments of pupils in their classes and asked to respond to what they saw in relation to their assessment of Speaking and Listening. The participating teachers were also asked to do a audio-tape assessment diary over a period of two weeks. They were asked to record any observations and thoughts which they considered relevant.

6. Data processing and analysis

All the interviews and diaries were transcribed. The transcriptions were then segmented into meaning units⁴ (Hitchcock and Hughes, 1989). Given the semi-structured nature of the interviews a great deal of the teacher discourse centred around the topics raised by the interviewer. This in-built patterning made the initial segmentation of the data a more ordered process than it might have been otherwise. The corpus was then encoded and imported into the software package NUD-IST for detailed investigation and analysis. In the segmentation stage, indexing categories were established for each of the three interviews and audio diaries. This indexing structure was repeated for each informant. The indexing system was constantly revised during data coding.

⁴ The term 'meaning unit' is used here in a broad sense which is consistent with a number of similar terms in the literature, eg. Krippendorff's (1980) referential unit, Richards' (1985) proposition and Harris' (1994) propositional level meaning. It is also compatible with Säljö's (1988) conception of reality (within the phenomenographic tradition).

6.1 Description and assessment: scale meanings and teacher/assessor meanings

6.1.1 Level 1: Level Description and teacher comments

In this section of the paper we examine some teacher comments about aspects of Level 1 Description and aspects which they report as missing from the Level Description. These comments were not responses to direct questions focused on Level Descriptions; they were offered as remarks apposite to the unfolding discussion at hand at various points of the interviews.

As part of the first interview phase (Espoused Theory), teachers were asked to comment on the way in which particular criteria were described in each of the first four Level Descriptions.⁵ Confidence, for instance, is specifically mentioned at Levels 2, 3 and 4 but not at Level 1. Where informants made a case for criteria to be included in a specific level these were coded and then analysed. All other instances in the interview data in which informants differentiated between two levels or defined a level, either generally or in relation to a particular child, were also analysed together with the data from the Espoused Theory interviews. Eight sites related to Level 1 Descriptions were identified in the discourse data. The eight sites are drawn from the data of three informants. The Phase 1 data for the fourth informant contained no mentions of categories missing in Level 1 and the Phase 2 and 3 data for this informant contained no relevant mentions of the Level Descriptions.

The Level 1 Description reads:

Pupils talk about matters of immediate interest. They listen to others and usually respond appropriately. They convey simple meanings to a range of listeners, speaking audibly, and begin to extend their ideas or accounts by providing some detail.

Table 1 below shows the relationship between the published Level 1 criteria and the accounts of individual teachers. Table 2 (following)

⁵Pupils at the end of Key Stage 1 (7-year-olds) are assessed with reference to the Levels 1, 2, 3 and 4 in the National Curriculum English.

indicates segments of teacher discourse which cannot be reconciled with Level 1 criteria.

The first thing which is striking about Table 1 is how little mention there is of Level 1 criteria. Interpretation of the degree of convergence between teachers' accounts of what constitutes a Level 1 and the account provided by the National Curriculum Level Descriptions is here inhibited by the limited number of references to Level 1 behaviour in the data. However, the fact that criteria which are not related to those described in the published scales are being used by teachers is suggestive (see Table 2). A similar pattern of convergence and divergence emerges for the other levels. Table 1 contains an endorsement by Teacher 2 of the meaningfulness of the Level 1 statement, "Pupils talk about matters of immediate interest".

NC Level 1 Descriptions	Informant comments which are broadly consistent with NC description
"Pupils talk about matters of immediate interest"	455 T2: and she will talk in much more structured language then. She
	456 still flounders for the appropriate words, though. But it
	457 shows that she can talk when she wants to, but she has to
	458 feel very strongly about it to do it, which sort of upholds
	459 level one, they talk about matters of immediate interest
"They listen to others and usually respond appropriately"	No data
"They convey simple meanings to a range of listeners,...	No data
...speak audibly...	No data
...and begin to extend their ideas or accounts by providing some detail"	No data

Table 1. NC Level 1 Descriptions and related comments which are broadly consistent with NC description

(Note: Informant extract in column 2, row 2 is repeated in Table 2)

However, the association with emergent language (*"flounders for the appropriate words"*) and with the teacher's inferences about the conditions for the child to talk (*"she has to feel very strongly about*

it to do it") suggest that highly interpretative processes are involved in arriving at a judgement.

Other Dimensions relevant to Level 1	Informant discourse
1. Confidence	<p>a 427 T4 : but generally at level one I'm looking to 428 develop confidence and I wouldn't be picking too much 429 on mistakes that I might want to improve on. That 430 would come later</p> <p>b 440 INT: For instance, at level 1. [Confidence is] not actually mentioned in 441 the scales at level 1, but is it an important criteria for you? 442 T3: For me it is, yes. It shows that they want to communicate and 443 are trying to communicate. It really is quite important. So, on a 444 level 1 basis 445 INT: It doesn't appear in the scales. 446 T3: It doesn't, no. I should have brought the records because on 447 ours we've got something like, they go off and can do a simple 448 job. In other words, we send them off to ask somebody something 449 and another member of staff or something to go and collect something</p> <p>c 710 INT: But they're certainly putting something across. 711 T4 : Well, they are. And the willingness to do it, I 712 think, is a landmark. You know, that the level ones on 713 the whole wouldn't be prepared to but the level twos, 714 the confidence and the sort of growing vocabulary and 715 the sort of just general knowledge of the way the 71 languages are made up ().</p>
2. Silence; amount and nature of interaction	<p>a 831 T4: In the Hertfordshire criteria it talks about, it 832 distinguishes between level one, you might just have 833 the child talking to you and you might get answers back 834 but they might still be fairly silent really within 835 their peer group, or they might just, they're not 836 really involved in some sort of group activity, but at 837 level two they're looking for group interchange so they</p>
3. Standard English and vocabulary	<p>a 978 T4: you're talking specifically about standard English and 979 the way that things should be said, and you're talking 980 about vocabulary that could be or should be used. You 981 know, it's quite specific, whereas at level one it's 982 mainly coming from them and you're encouraging it, at 983 level two you're putting a bit in, but at level three I</p>
4. Grammar	<p>a 84 T3: she gets tangled up, she gets very tangled up in 85 her tenses when she speaks. 86 INT: This is Noshin?</p>

-
- 87 T3: Yes, em, and I think Assiya is similar.
 88 INT: Would that for you be a feature of a level 1 child, rather
 89 than a level 2?
 90 T3: Yes, yes.
- b 237 T3: So for example, she's said 'Why's he not writing hers book?'. So
 238 it's an example of about what I was saying about the pronoun.
 239 So, I mean, this is obviously a level 1. 'Mines isn't all down',
 240 so again moving towards, moving there towards level 2.
- c 455 T2: and she will talk in much more structured language then. She
 456 still flounders for the appropriate words, though. But it
 457 show that she can talk when she wants to, but she has to
 458 feel very strongly about it to do it, which sort of upholds
 459 level one, they talk about matters of immediate interest.
-

Table 2. Categories not contained in NC Level 1

(Note: Informant comment 4c extract is repeated from Table 1)

Table 2 contains teacher criteria which diverge from those in the published scale descriptions. Two teachers report that "Confidence" (Table 2: 1) is a meaningful criterion. Setting aside the difficult question of what Confidence might comprise and how degrees or different types of it might be described and identified at different levels, it is clear that, for some teachers at least, it constitutes a meaningful category. One teacher, with reference to "The Hertfordshire scales",⁶ (Table 2: 2 "Silence; amount and nature of interaction") contrasts the silence of a Level 1 child with the emergence of social language in groups at Level 2. Neither of these features are explicitly marked in the National Curriculum scales. Teacher 4 refers to "Standard English" (Table 2: 3) and contrasts its emergence in Level 3 with Level 1 where "it's mainly coming from them". The contrast here is presumably between input for and expectation of increasing awareness of Standard English and its use (Level 3) and a lack of formal language based teaching and outcomes for bilinguals at Level 1. Table 2:4, "Grammar", picks up a recurrent theme in the data which will be discussed later. This concerns sentence level grammar, word order and is variously described by informants in terms such as "muddling up sentences", "getting tenses

⁶ The Hertfordshire scales operate like an extended version of the National Curriculum scales with examples or vignettes of the types of behaviour which are relevant to each level

right" and "whether they're getting their grammar right". These concerns are focussed mainly on Level 1 and on bilingual children. The National Curriculum Level Descriptions have nothing to say on this. The notion of Standard English appears for the first time at Level 3. Standard English in the Level Descriptions is conceptualised as a universal requirement but there is no reference as to how this may relate to the emergent language of the bilingual pupils.

In terms of the total number of indexing units in the data, the informant data in Table 1 and Table 2 represent a minor and numerically small subset of data. Modest though they are, they serve as entry to discussion of a number of thematic groupings which cluster in the data. Of particular interest is the discourse associated with Grammar, particularly for bilingual children at Level 1 and Level 2, a topic we return to later in the paper.

Taken together, Table 1 and Table 2 suggest that teachers operate criteria both inside and outside of those published in the National Curriculum. Precisely how these are calibrated to level and related to specific pupil behaviours remains, however, unresolved.

6.1.2 The lack of construct definition, variability in the discourse and the role of teachers' professional knowledge

Table 1 and Table 2 give the impression that there is relatively little mention of National Curriculum criteria in the data and that other categories dominate. However, this is not so. National Curriculum categories and concepts do appear in teachers' talk but are not necessarily precisely described or necessarily tied to particular level, instead serving as codified and yet flexible discursual resources for describing, amongst other things, teachers' experiences of assessment and teaching and the attainments and behaviours of the children they teach. The data, in general, do not suggest that teachers eschew the National Curriculum categories, but rather that they talk about them in ways which are both intra and inter-subjectively variable. In accounting for this variability, we are not in a position to propose a firm model since we believe that where accounts of mental representation have been discussed, whether from a social or cognitive perspective, (Fransella 1984: Marton 1981: Moscovici 1984: Potter and Wetherall 1987: Woods 1996) they are either too restrictive in their focus to be explanatory

of data of the type we are dealing with or methodologically unsuited to the particular questions we are researching. We therefore make no apologies for locating ourselves generally within a Schutzian (Schutz 1970) framework and within the broader perspective of Berger and Luckmann's (1967) work.

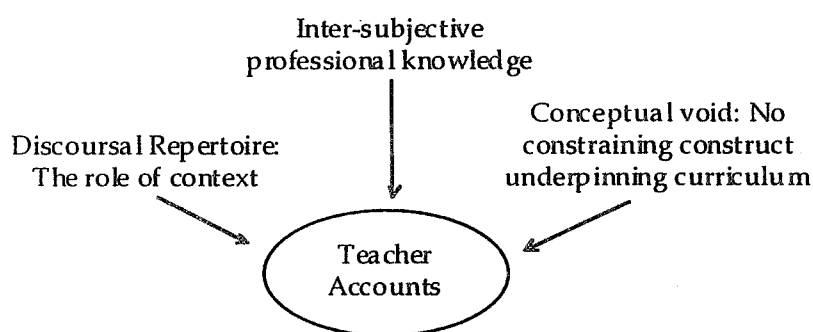


Figure 1. Simplified view of inputs to teacher accounts

We therefore propose an interim account of our data which makes sense of our findings to date. We suggest that there are at least three factors in interaction in the teachers' accounts. These we identify as: the conceptual space associated with the absence of any authoritative and constraining definition of the construct underlying the curriculum; the professional knowledge of teachers which we hold to be to some extent inter-subjectively shared; and naturally occurring contextual variation in the discourse. In our view these factors may be interacting to give a multi-faceted view of elements which might, other things being equal, be expected to be more unidimensional and internally coherent (see Figure 1. Simplified view of inputs to teacher accounts for a tentative graphical display of these processes).⁷ We suggest also that a compensatory mechanism may be observed. Where underlying construct and associated mechanisms lie unaccounted for in an apparently unified pedagogic and assessment system such as the National Curriculum,

⁷ As part of the triangulation of the data, a text analysis of the informant discourse with reference to the works of functional grammarians and critical discourse analysts (e.g. Eggins and Martin 1997; Fairclough 1989 and Harris 1994) was conducted. A particular focus on the use of clause level syntax, tense and modality was adopted. This assisted the interpretation of informant belief, commitment and truth value in the utterance.

then existing professional knowledge and teacher impressions (Eraut 1994) will serve to take their place. Elsewhere, we argue that initial training for teachers needs to take account more fully of propositional knowledge (especially in terms of language and language use)⁸ if the needs of bilingual children are to be adequately met (Leung and Teasdale forthcoming). Such training, in our view, would also serve to inform assessment practices. However, this should not be seen a substitute for consistency and clarity in relation to understanding and interpreting constructs.

6.1.3 Making sense of teachers' discourse and interpreting teacher perspective

As noted above, the informant data suggest that many of the National Curriculum categories form part of teachers' discoursal repertoires⁹ and come readily to them in talking about assessment, teaching and the attainment of individual children. Teacher 2 for instance talks about a child in these terms:

Extract 1

- 427 T2: ...she realises that there are more formal languages
 428 necessary. I think that's level 4, I think, no I'm
 429 getting confused. She's got the vocabulary at hand that she
 430 needs, whatever she's talking about. She's able to make
 431 points of view on a simple issue, not on a more complicated
 432 issue. But she is able to back up her point of view

Level 3 states:

⁸ There is currently no provision for pre-service training of ESL support teachers in Britain. Nor is there any standard component in teacher preparation courses for mainstream teachers which deals with principles and practice for dealing with the multilingual classroom.

⁹ Discoursal Repertoire is used here in a wider sense than Potter and Wetherell's (1987) use of the term. Our analysis, so far, does not suggest that distinctive and well-defined repertoires such as those identified by Potter and Wetherall in their studies account for our data in any meaningful way.

Pupils talk and listen confidently in different contexts, exploring and communicating ideas. In discussion, they show understanding of the main points. Through relevant comments and questions, they show they have listened carefully. They begin to adapt what they say to the needs of the listener, varying the use of vocabulary and the level of detail. They are beginning to be aware of standard English and when it is used

Tables 1 and 2 and the segment from Teacher 2 highlight some of the problems of attempting any precision in using teachers' discourse to map and calibrate assessment instruments since many of the categories identified are not in any Level Description. The tables suggest that teachers use a wider range of concepts and categories in arriving at level decisions than are contained in the National Curriculum descriptions. One problem concerns the "angle of approach" when specific levels are discussed. When teachers explain the assessment of a child in terms of criteria not yet met, rather than criteria met and surpassed, the lack of a solid and explicitly marked floor can make for difficulties of interpretation. Thus, in the Teacher 2 extract "complicated" is contrasted with "simple". Reconciling these is hard since most scales require some middle ground. In the context of the National Curriculum the description for Level 3 states that, "Pupils talk and listen confidently in different contexts, exploring and communicating ideas... adapt[ing] what they say to the needs of the listener, varying the use of vocabulary and the level of detail". This suggests a higher level of competence than is expressed in the teacher's characterisation of "able to make points of view on a simple issue but not on a more complicated one".

A second problem relates to teacher perspective and the way in which episodes in the discourse data, whether foregrounded by interviewer identification of particular children or not, frequently take a particular child or group of children as a thematic unit and reference point. This class of data contains extensive commentary on a range of issues which teachers clearly see as being of importance in their daily assessment and teaching work. It includes a focus on concerns such as the progress a child has made, the age of particular child relative to the rest of the peer group, the emotional state of the child and any home, cultural and linguistic factors which affect

the child's performance. In relation to an English as Mother Tongue child, Teacher 3 comments that:

Extract 2

134 T3: *Yes, she's the middle child of 5. The mother has just had*
 135 *another child. And, she's good at, good listening skills and is*
 136 *articulate, you know quite a good vocabulary, she's got very,*
 137 *good supportive parents, but they're obviously very busy with all*
 138 *the children at home. And she can express herself quite well but*

The home circumstances of bilingual children also appear regularly in our sample of teacher assessment talk. The example which follows combines the two analysis categories, Home and Personality. Home is associated with 36 sites in the data and Personality with a further 47 sites, making both significant indexing categories in the data set:

Extract 3

535 T4: *She's one of the youngest in another big family. We've had all*
 536 *the family in school, a lot of English will be spoken in the*
 537 *home, mum's English is quite good, dad's isn't bad. She's another*
 538 *one of these who's moody and a bit difficult. I don't think you*
 539 *always see what she can do in any way and she's a bit of a*
 540 *manipulator, so will get other people to do things for her*
 541 *(indistinct), she'll play the system basically, but I think*
 542 *probably I would say about level 2, I don't think she's*
 543 *outstanding, I don't think her vocabulary is particularly wide*

The above sequences are fairly typical of the way in which teachers weave children's wider circumstances and attributes into talk about their attainments. Notable in the second extract is the co-location in the discourse of features related to home circumstances, personality and assessment level, suggesting that for teachers these elements may be integrally related. It is difficult to see how such a

dimension, however powerfully articulated by teachers, could be incorporated into assessment for accountability, since it would render all assessments contingent upon evaluation of other conditions, which may in many circumstances remain unknowable. However, in a formative view of assessment-as-pedagogy, such factors crucially affect provision, and recognition of them acts as essential contextual information by which assessments can be interpreted and acted upon in curriculum planning.

7. Teachers' perceptions of the published criteria: the special case of bilingual children

In an empirical study, Leung and Teasdale (1996) suggest that English primary school teachers reference their assessments of bilingual children to a Native Speaker model of language and language use. The National Curriculum Level Descriptions encourage such a position as they do not refer to bilingual pupils nor are there any statutory instruments concerned specifically with the English language attainments of bilingual children.

In the current study, we have grouped data sites in the discourse which are related to:

- A Bilingual and Mother Tongue children
(24 sites; with mentions by 4 teachers)
- B Accuracy and sentence level production by pupils
(47 sites; 4 teachers)
- C Notions of Developmental Sequence
(24 sites; 3 teachers).

Some of the data are coded at more than one site (this is particularly so for category A and category B. A was addressed by interview protocols, although unsolicited data are also included. B and C are represented by codings of data which arose unsolicited. These sites all bear directly upon issues of the relevance of the Level Descriptions to the measurement of the language attainments of bilingual pupils and their utility as information to plan provision for this substantial minority of children in British schools.

Level 3 is identified as a critical level for bilinguals by 3 out of the 4 teachers. Teacher 4 comments:

Extract 4

- 1021 *But for the bilinguals that's where the*
 1022 *language gets difficult because you're talking in terms*
 1023 *of quite complicated sentence structures or you're*
 1024 *using specific vocabulary, or vocabulary that can be*
 1025 *used in different ways*

For Teacher 2, Level 3 is a point of convergence for bilingual and English as Mother Tongue pupils as well as a major hurdle for bilinguals:

Extract 5

- 606 *I think it's by Level 3 it's nearly evening*
 607 *out because if they've reached Level 3, they're*
 608 *quite proficient in language*

Extract 6

- 835 *And so*
 836 *they're stuck at Level 2 for a very long time. It*
 837 *would be true of all the children. There's one or*
 838 *two that will () through and they still find there's*
 839 *a tremendous gap, that they've never met "bicycle" or*
 840 *never known "spokes"*

The National Curriculum does not attend to the issue of the leap from Level 2 to Level 3 for bilingual pupils, nor does it acknowledge what the teachers in our study report about the importance of sentence level grammar for bilingual learners around Level 1. This lack of linguistic description, even at low levels of analysis, goes back to a failure to attend to construct interpretation. Ultimately, this is an issue of fairness, involving consequential validity, since

assessment in the way it is located in the curriculum of the English state school system feeds into both teaching and accountability (Gipps, 1994). A failure to account for the circumstances of a significant minority of pupils by legislating a curriculum which takes no account of their needs or their distinctive uses of English, is indeed a fundamental threat to validity.

Assessment of pupils who are working towards, through, and out of Level 1 is identified by all 4 teachers as another problematic area. Discussion of bilingual learners at this level has a high co-occurrence in the data with discussion of pupil output in terms of accuracy and sentence level grammar:

Extract 7

79 T4: *I think this shows that he still finds it very*
80 *difficult to accept proper structures of English, still operating*
81 *round about Level 1*

For Teacher 2 the concern for accuracy as an indicator applies to both bilinguals and English as a Mother tongue children, although there is tacit acceptance that inaccuracy in this area is less usual for the latter group:

Extract 8

200 T2: *...if they weren't speaking*
201 *in proper sentences I would be worried, probably more so if*
202 *they were an English as first language ()*

A systematic differentiation between language knowledge at Level 1 and Level 2 appears to be operationalised in Teacher 3's audio-taped diary of her assessment practice:

Extract 9

61 *She used the verb 'aten'. She obviously realised it was the verb*
62 *to eat and she knows that 'ate' is related to this verb, but she's*
63 *applying her knowledge wrongly of past tenses. So this I would*
64 *assess as Level 2, as she is applying her knowledge of grammar rules*

8. Can teacher representations and construct representation be aligned?

The teacher data also contain evidence of teacher views on developmental sequence in language learning, although these data tend not to form a coherent whole nor necessarily to exhibit internal coherence. Three sequences from Teacher 3 are shown below:

Extract 10

144 T3: *...we're at a very*
145 *interesting stage with them now...*
146 *....they are beginning to offer their own*
147 *observations and they have the confidence to verbalise them*

Extract 11

247 T3: *it just makes you aware of the fact that she's*
248 *got onto the next sort of section of deciphering the*
249 *language. She is aware that pronouns have got to be used*
250 *and she is now trying to sort out, putting them in the right*
251 *order*

Extract 12

403 T3: *Yes () is the same there is about 6 or 7 of them*
404 *at that sort of stage, still de-scrambling, still trying to*
405 *sort out what they're going to say. Being aware, becoming*
406 *aware of the structures of the English language and*
407 *trying to put them correctly, so again this use of*
408 *pronouns, she was mixing them up. Putting them in the wrong*
409 *order. And then, just general use of some words she was using*
410 *inappropriate words.*

* * * * *

Teacher 3 seems to operate a notion of stages for her own assessment of children. It remains unclear exactly how these stages relate to each other (no attempt was made to explore this in the interviews) and precisely how they articulate to the National Curriculum Level Descriptions (again, no data were specifically elicited in this area). However, it seems from these and other references to stages, that for Teacher 3 the notion of developmental stages does represent an explanatory device which informs her assessment and teaching. Currently, there is very little information on teachers' metaphors and representations of their pupils' language learning. Unifying these so that teachers and educationalists working within the same system speak and understand common assessment of language is clearly an urgent need and one which is a prerequisite if a common set of constructs in use is to emerge. As Säljö (1988) comments:

If our research has managed to discern the conceptions of the phenomenon being investigated, we have—metaphorically speaking—a map of a territory in terms of which we can interpret how people conceive of reality. (p.44)

However, even if the will and the resources were available to influence and attempt to unify teachers' conceptions, a simple social engineering approach of tightening up the underlying construct and embarking on massive training and standardisation may not, on its own, work. Where professional cultures are so deeply ingrained as in English primary school teaching we need to understand these maps more fully in order to know from where to start.

Two points illustrate this. One of our teachers described her assessment practice:

Extract 13

98 INT: *This is very interesting. So when you're working with the*

99 *children, you observe and*

100 T1: *I absorb*

101 INT: *And then remember the information. Then when do you do the*

102 *sheets*

103 T1: *Every term*

-
- 104 INT: *So, it's like building up, day by day, week by week*
- 105 T1: *Yeah. I mean I can remember, [name of child] can't do this or finds*
- 106 *difficulty with this because we've done it so often in our*
- 107 *circle time that I can just remember it, each child individually.*
- 108 INT: *So, at the end of each term you sit down and*
- 109 *get all the sheets out and when you go down the list you see*
- 110 *the children's names and then you refer to the Level*
- 111 *Description?*
- 112 T1: *Yeah, I'm afraid so.*
- 113 INT: *There's no right or wrong.*
- 114 T1: *I only did it at that time. I don't think about this any other*
- 115 *time. I just do my usual thing.*

Clearly, for this teacher the National Curriculum scales have little usefulness and little meaning in terms of her practice. The teacher prefers to rely on her own resources for assessment. For another teacher, the National Curriculum assessment also offers a less useful system than that built up over time from her own experience:

Extract 14

- 1140 T: *Well, I think the things like, you know, are they*
- 1141 *using one word, or three words, or are they able to use*
- 1142 *a sentence. You know, are they prepared to talk to me.*
- 1143 *Are they happy to talk to their friends, or are they*
- 1144 *standing in the playground silently, you know, when*
- 1145 *everybody else is chattering away. You know, that sort*
- 1146 *of thing really, I look for that.*
- 1147 INT: *You don't (do) that specifically?*
- 1148 T: *Well, I'm far more interested in that than this,*
- 1149 *but the fact that it's not in here doesn't worry me*
- * * * * *

-
- 1150 *because I'm looking for it anyway. And I started to*
1151 *teach twenty years ago so, you know, the National*
1152 *Curriculum is an additional extra to me. It's not*
1153 *fundamental to what I'm doing, or at least, I shouldn't*
1154 *say that, but it is fundamental to a certain extent,*
1155 *but I've got my own patterns from way back that I'm*
1156 *still firmly versed in.*

It is possible, and we believe persuasive, to characterise assessment in the National Curriculum as operating through the intersubjectively negotiated agreements of a whole series of assessment users. Teachers are key players in this process. There is a need, therefore, to examine teachers' and assessors' "own patterns", to examine the ways that they work, to assess the degree to which intersubjective agreement amongst teachers exists and eventually to use these as one aspect in the development of scales.

Ultimately, all rated assessment depends on a convergence of understanding between the construct at the core of the assessment and teachers' representations of the construct. The precise degree to which teachers focus on the same things when observing the construct operationalised in performance is ultimately unknowable. High inter-rater reliability coefficients are certainly a source of comfort in this area. However, they are not the end of the story, since we need further evidence that raters are indeed attending to the same features. The accessing of teacher report data is not a easy way to get this information, nor can we claim with any certainty that it accesses teachers' representations and assessment decision making. However, insofar as raters and not the descriptions themselves can be conceptualised as the tool which effects the measurement instrument in rated performance, it is prudent to proceed cautiously with the assumptions we make.

9. Conclusion

In this paper we have reported some initial findings in a larger study which is aimed at developing an understanding of the way in which teacher representations affect both pedagogy and assessment. Whilst our focus is strictly contextualised to the English school

setting, we believe that the study does raise wider questions, particularly in relation to:

- the relationship between construct, curriculum and assessment
- the degree to which it can be taken on trust that raters conceptualise construct and its attendant universe of content in the same way
- curricular and assessment validity where distinctive and arguably non-homogeneous populations of pupils are concerned.

In the earlier sections of the paper, we illustrated the degree to which raters refer in their talk about assessment to categories outside the National Curriculum Level Descriptions as well as to those within them. We note, however, that their use of general categories which appear in the National Curriculum is not always consistent with the actual formulation which appears in the published scales. Variability and inconsistency in the way the teacher informants talk about particular phenomena at different times present a problem for data interpretation. Following Säljö (see footnote 2), we regard this as not untypical of conversational accounts.

Our data also suggest a tendency for teachers to talk about assessment from a whole-child and learning-needs perspective. Categories such as home circumstances and personality are important in the way they conceptualise the attainments and learning needs of particular pupils. Finally, the discourse data suggest that teachers use some categories such as silent period, size of vocabulary and sentence grammar differently for bilingual and monolingual pupils.

References

- Argyris, C. and D.A. Schön 1974 *Theory in Practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass
- Berger P. and T. Luckmann 1967 *The Social Construction of Reality: a treatise in the sociology of knowledge*. London: Allen Lane

- Eggins, S. & J.R. Martin 1997 'Genres and registers of discourse'. In Van Dijk, T.A. (ed.) *Discourse as Structure and Process*. London: Sage
- Eraut, M. 1994 *Developing Professional Knowledge and Competence*. London: Falmer Press
- Fairclough, N. 1989 *Language and Power*. London: Longman
- Fransella F. 1984 'The Relationship between Kelly's constructs and Durkheim's representations' in R. M. Farr and Moscovici S. (eds) *Social Representations*. Cambridge: Cambridge University Press
- Gipps, C. 1994 *Beyond testing: towards a theory of educational assessment*. London: Falmer Press
- Gunn, M. 1995 'Criterion-based assessment: a classroom teacher's perspective' In G. Brindley (ed.) *Language Assessment in Action*. Macquarie University, NSW: National Centre for English Language Teaching and Research
- Harris, S. 1994 'Ideological exchanges in British magistrates courts.' In Gibbons, J. (ed) *Language and the Law*. London: Longman
- Hitchcock, G. & Hughes, D. 1989 *Research and the Teacher: a qualitative introduction to school-based research*. London: Routledge
- Krippendorff, K. 1980 *Content Analysis: an introduction to its methodology*. London: Sage
- Leung C. and A. Teasdale 1996 'What do teachers mean by speaking and listening: a contextualised study of assessment in the English National Curriculum.' Paper presented at the 18th Annual Language Testing Research Colloquium, Tampere.
- Leung C. and A. Teasdale (forthcoming) 'Initial Teacher Training and Professional Knowledge' in I. McGrath and H. Trappes-Lomax (eds) Title to be confirmed
- Marton F. 1981 'Phenomenography—Describing conceptions of the worlds around us.' *Instructional Science* 10 (177-200)
- * * * * *

-
- Messick, S. (1989) 'Validity.' In R. Linn (ed) *Educational Measurement* (3rd edn) American Council on Education, Washington: Macmillan
- Moscovici S. 1984 'The phenomenon of social representations' in R. M. Farr and S. Moscovici (eds) *Social Representations*. Cambridge: Cambridge University
- Potter J. and M. Wetherall 1987 *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: Sage Publications
- QSR 1996 NUD-IST. Victoria: LaTrobe University. (Distributed by Sage, London.)
- Säljö, R. 1988 'Learning in educational settings: methods of enquiry.' In P. Ramsden (ed.) *Improving Learning: new perspectives*. London: Kogan Pag.
- SCAA 1995 *English in the National Curriculum*. London: HMSO
- SCAA 1996 *Teaching EAL: a framework for policy*. London: SCAA
- Schön, D.A. 1987 *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass.
- Schutz, A. 1970 *On Phenomenology and Social Relations*. Chicago and London: University of Chicago Press
- Woods, D. 1996 *Teacher Cognition in Language Teaching*. Cambridge: Cambridge University Press

10. Appendix 1: National Curriculum English (1995) Attainment Targets: Speaking and Listening

10.1 Level 1

Pupils talk about matters of immediate interest. They listen to others and usually respond appropriately. They convey simple meanings to a range of listeners, speaking audibly, and begin to extend their ideas or accounts by providing some detail.

10.2 Level 2

Pupils begin to show confidence in talking and listening, particularly where the topics interest them. On occasions, they show awareness of the needs of the listener by including relevant detail. In developing and explaining their ideas they speak clearly and use a growing vocabulary. They usually listen carefully and respond with increasing appropriateness to what others say. They are beginning to be aware that in some situations a more formal vocabulary and the tone of voice are used.

10.3 Level 3

Pupils talk and listen confidently in different contexts, exploring and communicating ideas. In discussion, they show understanding of the main points. Through relevant comments and questions, they show they have listened carefully. They begin to adapt what they say to the needs of the listener, varying the use of vocabulary and the level of detail. They are beginning to be aware of standard English and when it is used.

10.4 Level 4

Pupils talk and listen with confidence in an increasing range of contexts. Their talk is adapted to the purpose: developing ideas thoughtfully, describing events and conveying their opinions clearly. In discussion, they listen carefully, making contributions and asking questions that are responsive to others' ideas and views. They use appropriately some of the features of standard English vocabulary and grammar.

N.B. The National Curriculum has eight levels; only the first four level descriptions (relevant to Key Stage 1, ages 5 to 7) are shown here.