

Comparative studies of mathematics teachers' observable learning objectives: validating low inference codes

Paul Andrews

Published online: 19 November 2008
© Springer Science + Business Media B.V. 2008

Abstract Videotape is an increasingly used tool in cross-national studies of mathematics teaching. However, the means by which videotaped lessons are coded and analysed remains an underdeveloped area with scholars adopting substantially different approaches to the task. In this paper we present an approach based on generic descriptors of mathematics learning objectives. Exploiting live observations in five European countries, the descriptors were developed in a bottom-up recursive manner for application to videotaped lessons from four of these countries, Belgium (Flanders), England, Hungary and Spain. The analyses showed not only that the descriptors were consistently operationalised but also that they facilitated the identification of both similarities and differences in the ways in which teachers conceptualise and present mathematics that resonated with the available literature. In so doing we make both methodological and theoretical contributions to comparative mathematics research in general and debates concerning the national mathematics teaching script in particular.

Keywords Comparative education research · English mathematics teaching · Flemish mathematics teaching · Generic codes · Hungarian mathematics teaching · Mathematics education · Observable learning outcomes · Spanish mathematics teaching · Teaching script · Video tape

1 Introduction

The mathematics education traditions of Europe (METE) project, funded by the European Union, examined the teaching of mathematics to students in the age range 10–14 in England, Finland, Flanders¹, Hungary and Spain; countries which represent well the geographical and cultural diversity of the continent. The project team spent the first year of

¹Throughout our study we have regarded Flanders, one of the two autonomous regions of Belgium, as a country in the manner of the third international mathematics and science study (TIMSS) and its repeats.

P. Andrews (✉)
Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge CB2 8PQ, UK
e-mail: pra23@cam.ac.uk

its 2-year programme developing, in a bottom-up iterative manner, a framework for analysing mathematics teachers' practice. During the second year the framework was used to analyse sequences of videotaped lessons taught on standard topics representative of all project countries' curricula. One section of the framework focused on the learning outcomes teachers were observed to privilege and it is on validating and demonstrating the effectiveness of this particular scale for comparative mathematics education research that this paper focuses.

2 Comparative mathematics education

Underpinning many comparative projects has been the belief that "studying teaching practices different from one's own can reveal taken-for-granted and hidden aspects of teaching" (Hiebert et al. 2003, p. 3) and many have unveiled unnoticed but culturally located teaching practices characteristic of the systems under scrutiny (Stigler, Gallimore & Hiebert 2000; Wilson, Andrew & Sourikova 2001). In particular, "comparing teaching and learning processes in different countries promises to help us to become more aware of our own implicit assumptions concerning the learning of mathematics" (Knipping 2003, p. 283). Moreover, TIMSS, PISA and their respective repeats have highlighted substantial variation in the mathematics achievement of students from economically comparable countries and provoked "a growing body of research aimed at unraveling the cultural and instructional factors that explain why children in certain countries so dramatically outperform their counterparts in other countries" (Yoshida, Verschaffel & De Corte 1997, p. 329). In particular, due to the success of a number of Pacific rim countries, many of these studies have compared aspects of mathematics education in either the US or England with China (including Hong Kong), Japan, Korea or Taiwan (An, Kulm & Wu 2004; Cai & Wang 2006; Leung 1995; Lin 1988; Ma 1999; Stigler & Perry 1990) and have identified substantial differences in the mathematics education traditions of what Tweed and Lehman (2002) have described as the Confucian and Socratic traditions.

However, to assume that European countries, with their shared Judeo-Christian cultural heritage, should evolve isomorphic educational systems would be naive as research has identified substantial culturally located differences in the educational traditions of the continent beyond the four dimensions proposed by Hofstede (1986). These include, for example, the influence of the Protestant and Catholic churches on the development of education in England and France respectively (Sharpe 1997) and societal emphases on the individual, community and nation underpinning educational expectations in England, Denmark and France respectively (Osborn 2004). Thus, it seems logical to accept that while cultural differences, no matter how small, impact in various ways on teachers' professional conduct, comparative research would be more likely to provoke systemic change if it were located in culturally similar rather than culturally dissimilar contexts.

A particular theme in recent comparative mathematics education research has been the explication of the national mathematics script. Its advocates, although there are critics (LeTendre, Baker, Akiba, Goesling & Wiseman 2001), argue that teachers in one country behave in ways that identify them more closely with their compatriots than teachers elsewhere (Schmidt et al. 1996). It embodies the pedagogical strategies which, through repeated enactment, are typical of a country's lessons and appear routine and beneath the consciousness of most teachers (Cogan & Schmidt 1999). Explanations for such patterns draw on beliefs that cultures "shape the classroom processes and teaching practices within

countries” (Knipping 2003, p. 282), to the extent that they are so “deep in the background of the schooling process... so taken-for-granted... as to be beneath mention” (Hufton & Elliott 2000, p. 117).

In this paper we discuss the development of the METE project’s coding schedule and an emergent imperative to identify and make operational essentially generic codes with the power to expose meaningful similarities and differences in the ways teachers present mathematics to their students. This is followed by analyses highlighting the sensitivity of the framework to cross-cultural similarities and differences in teachers’ actions. Finally, drawing on qualitative summaries of project lessons taught on percentages, we discuss the potential of the framework as a tool for comparative mathematics education research in general and an examination of the teaching script in particular.

3 The METE project

The mathematics education traditions of Europe project was a European Union-funded study of the teaching of standard mathematical topics to students in the age range 10–14 in Flanders, England, Finland, Hungary and Spain. In addition to the distinguishing politico-historical characteristics discussed above, the performance of project countries on TIMSS and PISA reflect well the range of European mathematics achievement. Moreover, they reflect not only Europe’s geographical diversity but also variation in factors like educational expenditure and demographics (OECD 2007).

The aim of the project was to examine, by means of both qualitative and quantitative approaches, how teachers structure students’ opportunities for learning. In this regard opportunity for learning was less a measure of curriculum coverage, as in the FIMS, than an analysis of the didactic strategies exploited and mathematical knowledge and skills encouraged by teachers. In so doing our intention was to make both methodological and theoretical contributions to comparative mathematics education research in general and the teaching script debate in particular. Thus, particularly in respect of the quantitative element of the study, decision making was informed by the desire to avoid the outcome of Stigler and Perry’s (1990) study of mathematics teaching in Japan, China and the US, whereby every examined dimension differentiated the systems under scrutiny.

4 Method

According to Givvin, Hiebert, Jacobs, Hollingworth and Gallimore (2005, p. 313) a “key to making progress on uncovering national or global patterns...is to examine classroom teaching directly” and the use of video cameras offers several advantages over traditional methods such as direct observation (Hiebert et al. 2003), although Ulewicz and Beatty (2001, p. 17) caution against an over-reliance on video tapes, arguing that they “cannot capture everything that happens in a classroom”. In respect of mathematics education, the TIMSS Video Studies (Stigler, Gonzales, Kawanaka, Knoll & Serrano 1999; Hiebert et al. 2003) and the Learner’s Perspectives Study (Clarke, Keitel & Shimizu 2006; Clarke, Emanuelsson, Jablonka & Mok 2006) have not only exploited video recordings to great effect but also informed considerably the development of our own work.

The main data set comprised four sequences of videotaped lessons from each country. Each sequence was taught by a teacher regarded locally as effective in the manner of the learner’s perspective study (Clarke 2002). Thus, generality, the goal of the TIMSS video

studies, was being traded for insights concerning national perceptions of best practice. Each sequence comprised four or five lessons on one of four standard topics; percentages and polygons in grades 5 or 6 and polygons again and linear equations in grades 7 or 8. These topics were representative not only of all project curricula but also the transition of school mathematics from concrete and inductive to abstract and deductive. Thus, between sixteen and twenty lessons per country were to be recorded and, apart from England where curricular recommendations² allowed for the recording of only fifteen, this was achieved.

In terms of data capture, a single camera was focused on all utterances made by the teacher and as much of the board-work as possible. In this manner, videographer decisions as to what should be the focus of attention were minimised (Ulewicz & Beatty 2001). In all cases, a tripod-mounted camera was placed at the side or rear of the room; teachers wore radio microphones while telescopic microphones captured as much student-talk as possible. Since our attention was on the ways in which teachers structured opportunities for learning, we felt that the integrity of data collection was not greatly compromised through lost student talk. After filming each videotape was compressed and transferred to CD ROM for copying and distribution, before being coded by its home team against the project's schedule.

Video-derived data are amenable not only to various forms of analysis but also present multiple opportunities for revisitation. In this regard the METE team's intention was to exploit this variety and several initial qualitative analyses have been reported (Carrillo & Climent 2005; Depaepe, De Corte, Op't Eynde & Verschaffel 2005; Sayers & Andrews 2005; Szeredi & Török 2005). However, project data were also amenable to quantitative analyses, particularly in respect of identifying replicable teacher behaviours and objectives. In this paper we present a summary of a bottom-up development of a coding schedule designed to address project research questions concerning teachers' conceptualisation and delivery of mathematics to students. In particular we discuss, from the perspective of demonstrating its methodological contribution to comparative mathematics education research, its potential for highlighting teachers' inferable learning objectives and their interactions.

The coding schedule was developed during the first year and although details of its development can be found elsewhere (Andrews 2007a) we offer here an extended summary. The initial intention was to exploit existing frameworks and Kilpatrick, Swafford and Findell's (2001) strands of *mathematical proficiency*, Hiebert and Lefevre's (1986) distinction between *conceptual knowledge* and *procedural knowledge* and Skemp's (1976) *relational* and *instrumental understanding* provided useful starting points. However, early discussions soon revealed that concepts assumed universal had contextualised meanings. This realisation led to a bottom-up process of live observation, discussion and code development for 1 week in each country. One lesson was observed each morning by at least one member of each project team. Each observation was supported by the quiet translation of a home colleague and the videotaping of the lesson for discussion the same afternoon. Discussions, which focused on identifying characteristic behaviours and activities suitable for coding, were supported by the playing back, as many times as was necessary, of the video-recording of the observed lesson.

Initial discussions were characterised by two difficulties. The first was a tendency to account for every conceivable form of teacher action and objective. The second, due to most team members being experienced mathematics teacher educators, concerned a

² In addition to an acknowledgement of a mandated national curriculum, English teachers are invited to work within a voluntary framework in which the number, and particular content, of lessons within a topic are specified.

collective inability “to suspend the desire to evaluate teachers’ efforts against unarticulated indicators of teaching quality” (Andrews 2007a, p. 496). Fortunately, over subsequent observation weeks, all of which were separated by team meetings at which the full project team discussed progress against further viewings of videotaped lessons, these problems were resolved. Firstly, as colleagues’ preconceptions were challenged and mediated “we became more skilled at making objective statements about what was seen, rather than what was missing” (Andrews *ibid*). Secondly, complex codes were rejected in favour of simple to activate descriptors focused on generic activities and objectives.

Importantly, all those later involved in the coding of video-taped lesson were involved in the cyclic process of code development. Moreover, observing lessons as they unfolded allowed colleagues to experience, and therefore understand, the cultural contexts in which they occurred. Thus, the project team felt confident not only that the outcome of the year’s work was a robust schedule that all colleagues could operationalise but also that the process ensured the content validity of the framework. That is, the active participation of colleagues, almost all of whom were experienced observers as teacher educators, classroom researchers or both, contributed significantly to the development of a schedule with high levels of applicability to the mathematics lessons of each project country.

A significant consequence of this process, deriving from a collective acceptance that all lessons are, in one way or another, episodic, was a decision to apply the codes to a generic episode. That is, for the purposes of data collection and analysis a generically defined episode, which was that part of a lesson where the teacher’s didactic intent remained constant, was accepted as the unit of analysis to which any of the project’s generic codes could be applied. Thus an episode could be short and managerial, as in the taking of a register or long and didactic, as in a lengthy period of seatwork. This generic episode, along with generic codes, provided a simple yet effective framework for describing and analysing project lessons.

The final version of the framework comprised three sections. The first, which is reported in this paper³, comprised seven items describing the observable learning outcomes of the episodes under scrutiny. The second comprised four items reflecting the conceptions of mathematics underpinning the tasks teachers present their students, while the third comprised ten items addressing teachers’ observable didactic strategies. The coding schedule’s descriptors were applied by members of each team to every episode of each of that country’s lessons as either present (1) or absent (0), with the consequence that few episodes were not multiply coded. Importantly, the affordance of multiple codings allowed for important distinctions in teacher behaviours to be highlighted. For example, an episode during which teachers presented an algorithm with no attempt to link it to underlying concepts would have been coded for procedural knowledge only. Where teachers drew on students’ prior understanding to facilitate their understanding of an algorithm an episode would have been coded for both procedural knowledge and conceptual knowledge. Working definitions of the seven generic learning outcomes can be seen in Table 1.

5 Results

In this paper we present several analyses to demonstrate how our inductively derived generic codes, when applied to generic episodes, facilitated meaningful comparisons. We also show how the interactions of the different objectives allude to differences, which may

³ This paper focuses on the one section of the schedule for no other reason than lack of space.

Table 1 Working definitions for each of the seven learning outcomes coded in project lessons

Conceptual knowledge (CK)	The teacher encourages the conceptual development of his or her students
Derived knowledge (DK)	The teacher encourages the development of new mathematical knowledge or entities from existing knowledge or entities
Structural knowledge (SK)	The teacher emphasises connections between different mathematical entities
Procedural knowledge (PK)	The teacher encourages the acquisition of skills, procedures, techniques or algorithms
Mathematical efficiency (ME)	The teacher encourages learners' development of procedural flexibility, awareness of elegance or critical comparison of working
Problem-solving (PS)	The teacher encourages learners' engagement with the solution of non-trivial or non-routine tasks
Mathematical reasoning (MR)	The teacher encourages learners' development and articulation of justification and argumentation

or may not be rooted in national traditions, in the coherence of the learning experiences offered students.

As indicated above, each country's team coded all the lessons videotaped in that country. In addition, the first two lessons in each sequence were transcribed and translated into English. This allowed for the production of subtitles, which enabled the English team to watch and code between 40% and 50% of each other country's lessons and determine inter-coder reliability. To this end the METE team exploited the Cohen Kappa coefficient, which accounts for agreements due to chance. Assuming a threshold value of 0.75, acceptable levels of inter-coder reliability were found between the coders of England and Flanders ($\kappa=0.877$), England and Hungary ($\kappa=0.875$) and England and Spain ($\kappa=0.793$). The first two figures fall well inside the "almost perfect" band of agreement, while the latter is essentially at the upper limit of the "substantial" band of agreement (Landis & Koch 1977, p. 165). Unfortunately, due to circumstances beyond the Finnish team's control, and despite its playing a full role in the development of the project schedule, the Finnish lessons have not been coded. Therefore this report is based on data from England, Flanders, Hungary and Spain.

5.1 A first level analysis of the observable learning objectives of project episodes

The mean of episodes coded for each observable outcome was calculated for each country. Presented in Table 2 they show that teachers regularly emphasise the conceptual and procedural knowledge of their students while derived knowledge is a rarity anywhere. This conformity of objective was supported by the analyses of variance (Table 3) showing none of these as influenced by nationality. It can also be seen that reasoning is a frequently observed outcome although, as indicated by the ANOVA, there were nationally located differences.

The remaining three outcomes, structural knowledge, mathematical efficiency and problem-solving were highlighted by the ANOVA as prone to the influence of nationality. To understand these differences we compared each country's data with those of all others—a project level comparison—and with those of each other country—a pair-wise comparison. In respect of the latter we had to consider whether multiple comparisons would require amendment to the significance levels. The Fisher least significant difference (LSD) requires no change to probability thresholds providing an analysis of variance has proved significant. Our view, despite a veneer of generality embedded in almost 400 episodes, was that our limited sample may have increased the risk of type I errors. Consequently, we were reluctant to accept the apparent liberalism of the LSD. Alternatively, the Bonferroni correction, which requires no initial ANOVA, recommends that typically acceptable

Table 2 The figures below show the percentage of each country's episodes in which each learning outcome was observed

	Flanders	England	Hungary	Spain	All
Conceptual knowledge	71.2	78.6	64.1	77.3	73
Derived knowledge	4.5	1.0	6.4	2.7	4
Structural knowledge	17.1 ^{ac}	1.0 ^{3, agk}	39.7 ^{5, cgm}	14.7 ^{km}	17
Procedural knowledge	56.8	53.4	51.3	68.0 ⁹	57
Mathematical efficiency	12.6 ^{1, d}	9.7 ^{4, h}	35.9 ^{6, dhn}	14.7 ⁿ	17
Problem-solving	7.2 ^{2, bef}	20.4 ^{b1}	30.8 ^{7, c}	38.7 ^{10, fl}	22
Reasoning	35.1	30.1 ^j	44.9 ^{8, jp}	25.3 ^p	34
Total episodes	111	103	78	75	367
Total lessons	20	15	18	16	69
Total didactic codes	299	237	320	264	1120
Codes per episode	2.7 ¹¹	2.3 ¹²	4.1 ¹³	3.5 ¹⁴	3.05

Numeric superscripts refer to significant project level differences as identified by Mann–Whitney *U*-tests. Alphabetic subscripts refer to significant pair-wise differences as identified by Mann–Whitney *U*-tests subsequent to a significant ANOVA. The adapted Bonferroni correction, as applied to the pair-wise differences, allows a threshold probability of $0.05/3=0.0167$

$$^1 z=2.123, p=0.034$$

$$^2 z=4.578, p=0.000$$

$$^3 z=5.078, p=0.000$$

$$^4 z=2.363, p=0.018$$

$$^5 z=7.081, p=0.000$$

$$^6 z=5.613, p=0.000$$

$$^7 z=2.010, p=0.044$$

$$^8 z=3.137, p=0.002$$

$$^9 z=2.374, p=0.018$$

$$^{10} z=3.799, p=0.000$$

$$^{11} z=3.320, p=0.001$$

$$^{12} z=7.071, p=0.000$$

$$^{13} z=7.583, p=0.000$$

$$^{14} z=3.968, p=0.000$$

$$^a z=3.625, p=0.000$$

$$^b z=2.808, p=0.005$$

$$^c z=4.465, p=0.000$$

$$^d z=4.490, p=0.000$$

$$^e z=4.241, p=0.000$$

$$^f z=5.258, p=0.000$$

$$^g z=7.170, p=0.000$$

$$^h z=4.604, p=0.000$$

$$^j z=2.547, p=0.011$$

$$^k z=3.588, p=0.000$$

$$^l z=2.672, p=0.008$$

$$^m z=3.912, p=0.000$$

$$^n z=3.311, p=0.001$$

$$^p z=2.981, p=0.003$$

Table 3 The results of Kruskal–Wallis analyses of variance showing the impact of nationality on the frequency of applied codes

	CK	DK	SK	PK	ME	PS	MR
χ^2	4.53	4.33	58.78	5.69	32.30	29.51	10.86
Df	3	3	3	3	3	3	3
P	0.209	0.228	0.000	0.128	0.000	0.000	0.013

probability thresholds should be divided by the number of comparisons being made. Thus, pair-wise comparisons across four countries yield six comparisons with a probability threshold $0.05/6=0.0083$. However, the conservative Bonferroni correction tends to increase the likelihood of type II errors. Our approach to this dilemma was to combine, in what we believe is a novel manner, the two conventions in the following way. Where analyses of variance were significant we would use a threshold probability of twice the Bonferroni value. Such a procedure insured against type I errors caused by a small sample and type II errors due to over-caution.

The two analyses, project level and pair-wise, yielded some interesting outcomes in terms of the frequencies with which structural knowledge, mathematical efficiency, problem solving and reasoning were observed. A significantly smaller proportion of English episodes was coded for structural knowledge at all pair-wise and project levels than elsewhere. The opposite was true of the Hungarian episodes with a significantly higher proportion of episodes coded for structural knowledge at the project and all pair-wise levels. Mathematical efficiency was also a reasonable discriminator with a significantly higher proportion of episodes being so coded in Hungary at the project and all pair-wise levels. Also, although there were no pair-wise differences apart from with Hungary, there were significantly smaller proportions of English and Flemish episodes coded for mathematical efficiency at their respective project levels. A lack of problem-solving in Flanders, at all pair-wise and project levels, distinguished the episodes of that country significantly from those of all others, while in England a significantly smaller proportion of episodes than in Spain was so coded. Indeed, when examined at the project level, there were significantly higher proportions of episodes coded for problem-solving in the Spanish and Hungarian episodes than elsewhere. Lastly, a significantly higher proportion of episodes was coded for reasoning in Hungary than elsewhere at the project and pair-wise levels with both England and Spain. We shall return to these findings later.

5.2 A second level analysis of the observable learning objectives of project episodes

The above results indicate both similarities and differences. The major similarities, that teachers emphasise their students' procedural knowledge and conceptual knowledge, seem unsurprising and offer some support for the validity of the instrument and the manner of its implementation. The major differences offer support for variation in the ways in which teachers from different countries present mathematics to their students. However, frequencies alone offer only a partial account and offer little in respect of explicating the complexity of mathematics teaching. In the following we offer a second level analysis which not only demonstrates how the generic codes interact in project episodes but also highlights substantial differences in the ways in which project teachers conceptualise and present mathematics to their students. In so doing we draw explicitly on the fact that if a learning outcome had been observed during an episode it would have been coded as present irrespective of which other strategies had already been coded. Thus, the majority of

episodes were multiply coded. This second analysis entailed the following. Whenever an episode was coded for one learning outcome, say *conceptual knowledge*, the record for that episode was checked for each of the remaining outcomes also. This process allowed us to count the number of episodes coded for *conceptual knowledge* with *procedural knowledge*, *conceptual knowledge* with *structural knowledge*, *conceptual knowledge* with *reasoning* and so on. In so doing, we were able, without analyses of variance, to examine some of the interactions between the learning outcomes in simple to report and accessible ways. Lastly, due to its infrequent occurrence, we have not examined episodes coded for derived knowledge with any other outcome, although it was retained as a second code. Included in each table are the results of chi square tests to determine the influence of nationality and Mann–Whitney *U*-tests, acknowledging when appropriate the adapted Bonferroni correction, highlighting both project level and pair-wise differences.

5.2.1 Episodes coded for conceptual knowledge and other outcomes

The figures of Table 4 show that the distributions of conceptual knowledge with derived knowledge, procedural knowledge and reasoning respectively were consistent across all countries' episodes. However, the distributions of conceptual knowledge with structural knowledge, mathematical efficiency and problem-solving were all influenced by nationality. At the pair-wise level a smaller proportion of English episodes was coded for conceptual knowledge with structural knowledge than from any other country, while a significantly higher proportion of episodes was so coded in Hungary. Both the Flemish and the English episodes comprised smaller proportions of episodes coded for conceptual knowledge with mathematical efficiency than in Hungary. The Flemish episodes also comprised a significantly lower proportion of episodes coded for conceptual knowledge with problem-solving than in either Hungary or Spain while the latter also comprised a higher proportion than in England.

At the project level Hungarian teachers taught a significantly higher proportion of episodes coded for conceptual knowledge with structural knowledge and mathematical efficiency respectively than teachers elsewhere whereas English teachers taught a smaller proportion of episodes coded for conceptual knowledge with structural knowledge than teachers elsewhere. Flemish teachers taught a significantly smaller proportion of episodes coded for conceptual knowledge with problem-solving while Spanish teachers taught a significantly higher proportion than teachers elsewhere.

5.2.2 Episodes coded for structural knowledge and other outcomes

The figures of Table 5 show that the distributions of episodes coded for structural knowledge with derived knowledge and procedural knowledge were consistent across the four countries' episodes. The pair-wise comparisons indicated only two significant results; higher proportions of Hungarian episodes were coded for structural knowledge with mathematical efficiency and problem-solving respectively than were the Flemish.

At the project level a smaller proportion of Hungarian episodes was coded for structural knowledge with conceptual knowledge than elsewhere, a smaller proportion of Flemish episodes was coded for structural knowledge with mathematical efficiency and problem-solving respectively than elsewhere, while a higher proportion of Hungarian episodes was coded than elsewhere. A higher proportion of Hungarian episodes was coded for structural knowledge with reasoning than elsewhere while the opposite was true for the Spanish episodes.

Table 4 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for conceptual knowledge

Derived knowledge ($p=0.119$)			
	0	1	
Flanders	71	5 (7)	76
England	80	1 (1)	81
Hungary	48	5 (9)	53
Spain	56	2 (3)	58
	255	13 (5)	268
Procedural knowledge ($p=0.132$)			
	0	1	
Flanders	38	38 (50)	76
England	40	41 (51)	81
Hungary	32	21 (40)	53
Spain	22	36 (62)	58
	132	136 (51)	268
Problem-solving ($p<0.0005$)			
	0	1	
Flanders ^{1, 2, a}	70	6 (8)	76
England ³	63	18 (22)	81
Hungary ¹	40	13 (25)	53
Spain ^{2, 3, b}	33	25 (43)	58
	206	62 (23)	268
Structural knowledge ($p<0.0005$)			
	0	1	
Flanders ^{4, 5}	61	15 (20)	76
England ^{4, 6, 7, c}	81	0 (0)	81
Hungary ^{5, 6, 8, d}	28	25 (47)	53
Spain ^{7, 8}	47	11 (19)	58
	217	51 (19)	268
Mathematical efficiency ($p<0.0005$)			
	0	1	
Flanders ⁹	68	8 (11)	76
England ¹⁰	72	9 (11)	81
Hungary ^{9, 10, e}	33	20 (38)	53
Spain	48	10 (17)	58
	221	47 (18)	268
Reasoning ($p=0.365$)			
	0	1	
Flanders	46	30 (39)	76
England	50	31 (38)	81
Hungary	27	26 (48)	53
Spain	39	19 (33)	58
	162	106 (39)	268

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U -tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z=2.61, p=0.009$

² $z=4.77, p<0.0005$

³ $z=2.62, p=0.009$

⁴ $z=4.19, p<0.0005$

⁵ $z=3.30, p=0.001$

⁶ $z=6.83, p<0.0005$

⁷ $z=4.07, p<0.0005$

⁸ $z=3.16, p=0.002$

⁹ $z=3.67, p<0.0005$

¹⁰ $z=3.65, p<0.0005$

^a $z=3.72, p<0.0005$

^b $z=4.07, p<0.0005$

^c $z=5.21, p<0.0005$

^d $z=5.82, p<0.0005$

^e $z=4.31, p<0.0005$

Table 5 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for structural knowledge

Conceptual knowledge ($p=0.017$)			
	0	1	
Flanders	1	15 (94)	16
England	1	0 (0)	1
Hungary ^a	9	25 (74)	34
Spain	0	11 (100)	11
	11	51 (82)	62
Procedural knowledge ($p=0.367$)			
	0	1	
Flanders	10	6 (38)	16
England	1	0 (0)	1
Hungary	17	17 (50)	34
Spain	4	7 (64)	11
	32	30 (48)	62
Problem-solving ($p=0.050$)			
	0	1	
Flanders ^{1, b}	16	0 (0)	16
England	1	0 (0)	1
Hungary ^{1, c}	23	11 (32)	34
Spain	9	2 (18)	11
	49	13 (21)	62
Derived knowledge ($p=0.738$)			
	0	1	
Flanders	15	1 (6)	16
England	1	0 (0)	1
Hungary	31	3 (9)	34
Spain	9	2 (18)	11
	56	6 (10)	62
Mathematical efficiency ($p=0.002$)			
	0	1	
Flanders ^{2, d}	15	1 (6)	16
England	1	0 (0)	1
Hungary ^{2, c}	16	18 (53)	34
Spain	10	1 (9)	11
	42	20 (32)	62
Mathematical reasoning ($p=0.049$)			
	0	1	
Flanders	7	9 (56)	16
England	1	0 (0)	1
Hungary ^f	10	24 (71)	34
Spain ^g	8	3 (27)	11
	26	36 (58)	62

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U -tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z=2.55$, $p=0.011$

² $z=3.14$, $p=0.002$

^a $z=1.97$, $p=0.049$

^b $z=2.37$, $p=0.018$

^c $z=2.41$, $p=0.016$

^d $z=2.56$, $p=0.010$

^e $z=3.81$, $p<0.0005$

^f $z=2.18$, $p=0.029$

^g $z=2.26$, $p=0.024$

5.2.3 Episodes coded for procedural knowledge and other outcomes

The figures of Table 6 show that the distributions of episodes coded for procedural knowledge with conceptual knowledge and derived knowledge respectively were consistent across all countries, while the remaining four outcomes were linked with 12 significant pair-wise differences. In detail, higher proportions of Hungarian episodes were coded for procedural knowledge with structural knowledge and mathematical efficiency respectively than those of any other country. Also, a higher proportion of Hungarian episodes was coded for procedural knowledge with reasoning than either the English or Spanish. By way of contrast, a smaller proportion of episodes were coded for procedural knowledge with structural knowledge in England than in any other country, while a smaller proportion of episodes was coded for procedural knowledge with problem-solving in Flanders than any other country.

At the project level Hungarian teachers taught a higher proportion of episodes coded for procedural knowledge with structural knowledge, mathematical efficiency, problem-solving and reasoning than teachers elsewhere. English teachers taught a smaller proportion of episodes coded for procedural knowledge with structural knowledge, Flemish teachers a smaller proportion of episodes coded for problem-solving and Spanish teachers, while teaching a higher proportion of episodes coded for problem-solving, taught a smaller proportion of episodes coded for reasoning than teachers elsewhere.

5.2.4 Episodes coded for mathematical efficiency and other outcomes

The figures of Table 7 show that the distribution of mathematical efficiency with all other codes bar structural knowledge was uniform across countries. In terms of pair-wise comparisons, a higher proportion of episodes was coded for mathematical efficiency and structural knowledge in Hungary than either England or Flanders. At the project level, similar results were obtained with a higher proportion of episodes coded for mathematical efficiency with structural knowledge than elsewhere in Hungary and a smaller proportion so coded in England than elsewhere.

5.3 Episodes coded for mathematical problem-solving and other outcomes

The figures of Table 8 show that the distributions of episodes coded for problem-solving with derived knowledge, procedural knowledge and mathematical efficiency respectively were uniform across countries. In respect of pair-wise differences a higher proportion of Hungarian episodes was coded for problem-solving with structural knowledge than in either England or Spain. Also, a higher proportion of Hungarian episodes was coded for problem-solving with reasoning than the English. At the project level a smaller proportion of Hungarian episodes was coded for problem-solving with conceptual knowledge, but a higher proportion of problem-solving with structural knowledge and reasoning respectively while a smaller proportion of English episodes was coded for structural knowledge than elsewhere.

5.3.1 Episodes coded for mathematical reasoning and other outcomes

The figures of Table 9 show that the distribution of episodes coded for mathematical reasoning with derived knowledge and procedural knowledge respectively were distributed uniformly across all four countries. As far as pair-wise differences were concerned, a

smaller proportion of Hungarian episodes was coded for mathematical reasoning with conceptual knowledge than in England or Spain, while a higher proportion of Hungarian episodes was coded for mathematical reasoning with structural knowledge than in Flanders, England or Spain. On the same theme, a higher proportion of Flemish episodes were so coded than in England. A higher proportion of episodes was coded for mathematical reasoning with mathematical efficiency in Hungary than in either Flanders or England. Lastly, a smaller proportion of episodes was coded for mathematical reasoning with problem-solving in Flanders than in either Hungary or Spain, while a smaller proportion of episodes were so coded in England than Spain.

At the project level Hungarian teachers taught higher proportions of episodes coded for structural knowledge, mathematical efficiency and problem-solving than teachers elsewhere. English teachers taught a smaller proportion of episodes coded for structural knowledge and mathematical efficiency than their overseas colleagues. Hungarian teachers taught a significantly smaller proportion and English teachers a significantly higher

Table 6 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for procedural knowledge

Conceptual knowledge ($p=0.058$)			
	0	1	
Flanders	22	38 (63)	60
England	14	41 (75)	55
Hungary	22	21 (49)	43
Spain	16	36 (69)	52
	74	136 (65)	210
Structural knowledge ($p<0.0005$)			
	0	1	
Flanders ^{1, 2}	54	6 (10)	60
England ^{1, 3, 4, a}	55	0 (0)	55
Hungary ^{2, 3, 5, b}	26	17 (40)	43
Spain ^{4, 5}	45	7 (13)	52
	180	30 (14)	210
Problem-solving ($p<0.0005$)			
	0	1	
Flanders ^{6, 7, 8, c}	58	2 (3)	60
England ⁶	43	12 (22)	55
Hungary ^{7, d}	28	15 (35)	43
Spain ^{8, e}	34	18 (35)	52
	163	47 (22)	210
Derived knowledge ($p=0.168$)			
	0	1	
Flanders	58	2 (3)	60
England	55	0 (0)	55
Hungary	43	0 (0)	43
Spain	52	0 (0)	52
	208	2 (1)	210
Mathematical efficiency ($p=0.001$)			
	0	1	
Flanders ⁹	52	8 (13)	60
England ¹⁰	48	7 (13)	55
Hungary ^{9, 10, 11, f}	25	18 (42)	43
Spain ¹¹	44	8 (15)	52
	169	41 (20)	210

Table 6 (continued)

Mathematical reasoning ($p=0.002$)			
	0	1	
Flanders	41	19 (32)	60
England ¹²	43	12 (22)	55
Hungary ^{12, 13, G}	22	21 (49)	43
Spain ^{13, H}	44	8 (15)	52
	150	60 (29)	210

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U -tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z=2.40, p=0.016$

² $z=3.53, p<0.0005$

³ $z=5.10, p<0.0005$

⁴ $z=2.80, p=0.005$

⁵ $z=2.90, p=0.004$

⁶ $z=3.02, p=0.003$

⁷ $z=4.23, p<0.0005$

⁸ $z=4.30, p<0.0005$

⁹ $z=3.27, p=0.001$

¹⁰ $z=3.27, p=0.001$

¹¹ $z=2.87, p=0.004$

¹² $z=4.30, p=0.005$

¹³ $z=3.51, p<0.0005$

^a $z=3.52, p<0.0005$

^b $z=5.29, p<0.0005$

^c $z=4.18, p<0.0005$

^d $z=2.20, p=0.028$

^e $z=2.43, p=0.015$

^f $z=4.13, p<0.0005$

^g $z=0.015, p=0.001$

^h $z=2.42, p=0.015$

proportion of episodes coded for conceptual knowledge than teachers elsewhere. Lastly, Spanish teachers taught a significantly higher proportion while Flemish teachers a significantly smaller proportion of episodes coded for problem-solving than elsewhere.

6 Discussion

In the following we attend to three different but not unrelated issues. The first concerns the extent to which the schedule satisfied the project team's objectives concerning simplicity of operationalisation but with the power to expose meaningful similarities and differences in the ways project teachers present mathematics to their students. The second, alluding to construct validity, concerns the extent to which the seven learning outcomes offers the potential for constructing models of teacher behaviour which others working in the field would recognise. The third, which resonates with notions of national teaching script,

Table 7 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for mathematical efficiency

Conceptual knowledge ($p=0.242$)			
	0	1	
Flanders	4	8 (67)	12
England	1	9 (90)	10
Hungary	10	20 (67)	30
Spain	1	10 (91)	11
	16	47 (75)	
47	63		60
Structural knowledge ($p<0.0005$)			
	0	1	
Flanders ¹	11	1 (8)	12
England ^{2, a}	10	0 (0)	10
Hungary ^{1, 2, b}	12	18 (60)	30
Spain	10	1 (9)	11
	43	20 (32)	63
Problem-solving ($p=0.54$)			
	0	1	
Flanders	9	3 (25)	12
England	7	3 (30)	10
Hungary	21	9 (30)	30
Spain	3	8 (73)	11
	40	23 (37)	63
Derived knowledge ($p=0.616$)			
	0	1	
Flanders	12	0 (0)	12
England	10	0 (0)	10
Hungary	28	2 (7)	30
Spain	10	1 (9)	11
3 (5)	63		
Procedural knowledge ($p=0.863$)			
	0	1	
Flanders	4	8 (67)	12
England	3	7 (70)	10
Hungary	12	18 60)	30
Spain	3	8 (73)	11
	22	41 (65)	63
Reasoning ($p=0.160$)			
	0	1	
Flanders	5	7 (58)	12
England	6	4 (40)	10
Hungary	9	21 (70)	30
Spain	7	4 (36)	11
	27	36 (57)	63

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U-tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z=3.00$, $p=0.009$

² $z=3.51$, $p<0.0005$

^a $z=2.33$, $p=0.020$

^b $z=4.56$, $p<0.0005$

Table 8 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for problem solving

Conceptual knowledge ($p=0.030$)			
	0	1	
Flanders	2	6 (75)	8
England	3	18 (86)	21
Hungary ^a	11	13 (54)	24
Spain	4	25 (86)	29
	20	62 (76)	82
Structural knowledge ($p<0.0005$)			
	0	1	
Flanders	8	0 (0)	8
England ^{1, b}	21	0 (0)	21
Hungary ^{1, 2, c}	13	11 (46)	24
Spain ²	27	2 (7)	29
	69	13 (16)	82
Efficiency ($p=0.335$)			
	0	1	
Flanders	5	3 (38)	8
England	18	3 (14)	21
Hungary	15	9 (38)	24
Spain	21	8 (28)	29
	59	23 (28)	82
Derived knowledge ($p=0.765$)			
	0	1	
Flanders	8	0 (0)	8
England	21	0 (0)	21
Hungary	23	1 (4)	24
Spain	28	1 (3)	29
	80	2 (2)	82
Procedural knowledge ($p=0.267$)			
	0	1	
Flanders	6	2 (25)	8
England	9	12 (57)	21
Hungary	9	15 (63)	24
Spain	11	18 (62)	29
	35	47 (57)	82
Mathematical reasoning ($p=0.014$)			
	0	1	
Flanders	6	2 (25)	8
England ³	15	6 (29)	21
Hungary ^{3, d}	7	17 (71)	24
Spain	18	11 (38)	29
	46	36 (44)	82

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U -tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z=3.53$, $p<0.0005$

² $z=3.25$, $p=0.001$

³ $z=2.80$, $p=0.005$

^a $z=2.89$, $p=0.004$

^b $z=2.29$, $p=0.022$

^c $z=4.75$, $p<0.0005$

^d $z=3.14$, $p=0.002$

concerns the potential contribution of such an instrument to comparative mathematics education research.

6.1 Instrument reliability

In respect of the first issue, the analyses, particularly the inter-coder reliability calculations, indicated that project colleagues, irrespective of their country of origin, were able not only to agree on the entry and exit points of lesson episodes but also the learning outcomes observed within them. This, following the difficulty experienced during the first year of the project, was a satisfactory outcome and suggests that the descriptors of both generic episodes and generic codes were straightforward to make operational. The analyses indicated, also, unsurprising similarities and, as we show below, explicable differences in teachers' subject-related emphases. Thus, the METE project's objectives appear to have been satisfactorily addressed.

Table 9 Below are the frequencies, with percentages in parentheses, of each learning outcome in relation to episodes coded for reasoning

Conceptual knowledge ($p=0.001$)			
	0	1	
Flanders	6	30 (83)	36
England ^{1, a}	0	31 (100)	31
Hungary ^{1, 2, b}	12	26 (68)	38
Spain ²	0	19 (100)	19
	18	106 (85)	124
Structural knowledge ($p<0.0005$)			
	0	1	
Flanders ^{3, 4}	27	9 (25)	36
England ^{3, 5, c}	31	0 (0)	31
Hungary ^{4, 5, 6, d}	14	24 (63)	38
Spain ⁶	16	3 (16)	19
	88	36 (29)	124
Mathematical efficiency ($p<0.0005$)			
	0	1	
Flanders ⁷	29	7 (19)	36
England ^{8, c}	27	4 (13)	31
Hungary ^{7, 8, f}	17	21 (55)	38
Spain	15	4 (21)	19
	88	36 (29)	124
Derived knowledge ($p=0.197$)			
	0	1	
Flanders	36	0 (0)	36
England	30	1 (3)	31
Hungary	34	4 (11)	38
Spain	18	1 (5)	19
	118	6 (5)	124
Procedural knowledge ($p=0.483$)			
	0	1	
Flanders	17	19 (53)	36
England	19	12 (39)	31
Hungary	17	21 (55)	38
Spain	11	8 (42)	19
	64	60 (48)	124

Table 9 (continued)

Problem-solving ($p < 0.0005$)			
	0	1	
Flanders ^{9, 10, g}	34	2 (6)	36
England ¹¹	25	6 (16)	31
Hungary ^{9, h}	21	17 (45)	38
Spain ^{10, 11, j}	8	11 (58)	19
	88	36 (29)	124

Included are the results of chi square tests (probabilities in brackets) indicating the extent to which distributions are influenced by nationality, and Mann–Whitney U -tests where numerical superscripts refer to significant pair-wise differences and alphabetic superscripts refer to significant project level differences.

¹ $z = 3.42, p = 0.001$

² $z = 2.72, p = 0.006$

³ $z = 2.97, p = 0.003$

⁴ $z = 3.28, p = 0.001$

⁵ $z = 5.44, p < 0.0005$

⁶ $z = 3.35, p = 0.001$

⁷ $z = 3.15, p = 0.002$

⁸ $z = 3.62, p < 0.0005$

⁹ $z = 3.83, p < 0.0005$

¹⁰ $z = 4.31, p < 0.0005$

¹¹ $z = 2.76, p = 0.006$

^a $z = 2.64, p = 0.008$

^b $z = 3.57, p < 0.0005$

^c $z = 4.10, p < 0.0005$

^d $z = 5.54, p < 0.0005$

^e $z = 2.28, p = 0.023$

^f $z = 4.26, p = 0.0005$

^g $z = 3.67, p < 0.0005$

^h $z = 2.55, p = 0.011$

^j $z = 3.00, p = 0.003$

6.2 Instrument validity and the teaching of percentages

In terms of the second issue, one approach to construct validation is to assess the extent to which the outcomes yielded by the instrument resonate with those derived from both other forms of project data (triangulation) and earlier studies. Thus, in the following we compare our analyses for each country against both the available literature and brief summaries of the qualitative analyses of the early lessons in each sequence on percentages. To facilitate this process we first consider the goals of percentage teaching. In relation to the Dutch tradition of realistic mathematics education, Van den Heuvel-Panhuizen (1994, p. 356) writes that “what percentages are should emerge in how the students are using them”... and “percentages are always related to something and that they therefore cannot be compared without taking into account to what they refer”. More generally, Depaepe et al.’s (2005) review of the literature concluded that the teaching of percentages should attend to procedural knowledge, conceptual knowledge and adaptive expertise. Hatano and Inagaki (1986) contrast routine expertise—knowledge and skills learnt with little understanding,

frequently by rote, and applied to familiar problems—with adaptive expertise—knowledge and skills which have sufficient and connected underlying understanding for the solution of non-familiar problems. Adaptive expertise, or the flexible application of knowledge, is an increasingly important objective in mathematics education (Baroody 2003; Kilpatrick et al. 2001) and requires an integrated and connected set of concepts and procedures. In respect of framing our discussion and acknowledging the high proportions of conceptual knowledge and procedural knowledge observed in project classrooms, the extent to which adaptive expertise is acknowledged will lie in the ways teachers promote, for example, structural knowledge and mathematical efficiency.

Also, by way of framing our discussion we introduce two concepts which emerged from our analyses and wider reading. The first concerns the *cognitive complexity* of the mathematics presented to learners and the second the *didactic coherence* of the manner of its delivery. Drawing on Schmidt et al.'s (1996) work, cognitive complexity reflects the different cognitive demands of different problem types (Hiebert & Wearne 1993; Stein, Grover & Henningsen 1996; Andrews 2003). Indeed, the “cognitive demands placed on a student—that is, the thinking a student is required to do, can be quite different depending on the type of question asked” (Schmidt et al. 1996, p. 81). However, cognitive complexity alone is insufficient to explain the ways in which teachers, or systems, structure opportunities for learning. The logic implicit in the sequencing of concepts, and the extent to which learners are offered a connected and integrated experience of mathematics, reflect some sense of the didactic coherence of the learner experience (Jennings & Dunne 1996; Kaiser, Hino & Knipping 2006).

6.2.1 Flanders

Flemish students appeared to experience fewer opportunities for problem-solving and mathematical efficiency than students elsewhere. These findings, in the light of earlier studies highlighting the difficulties faced by many Flemish students' when exposed to non-standard problems (Verschaffel, De Corte & Lasure 1994; De Bock, Verschaffel, Janssens, Van Dooren & Claes 2003; Van Dooren, De Bock, Depaep, Janssens & Verschaffel 2003), are probably not surprising. However, Flemish students experienced conceptual knowledge, structural knowledge, procedural knowledge and reasoning in frequencies commensurate, at both levels of analysis, with project norms. Such emphases, when viewed alongside the figures of Table 2 showing Flemish episodes attracting significantly fewer codes than the project norm, point towards a tightly and consistently defined Platonist conception of mathematics as a static but unified body of knowledge (Ernest 1989) or a coherent set of inter-related concepts (Thompson 1984). Thus, in respect of adaptive expertise, Flemish students appear to be presented with ample opportunities to acquire connected conceptual knowledge, but relatively few opportunities to exploit that knowledge.

This conclusion finds support in the qualitative analyses of the percentage lessons. The teacher, having invited students to bring to school any household items with some reference to the word percentage, began the first lesson by discussing what they thought the term meant and establishing a collective understanding concerning parts per hundred. This was followed, with the aid of base 10 number blocks that were gradually withdrawn over several lessons, by a series of problems which were solved collectively and which, through initially highlighting repeated addition, led her students to see the calculation of percentages as a process most effectively managed multiplicatively. Throughout a clear emphasis was maintained on learners' conceptual knowledge, focused on an emergent insistence that calculations were to entail “dividing the given amount by (one) hundred to calculate 1% of

that amount; and multiplying the result by the percentage” (Depaepe et al. 2005: 165). In so doing the teacher not only foregrounded the structural properties of percentages but also demonstrated a clear understanding of the logical sequencing of ideas embedded within the topic. However, despite her introductory activity, she did not return to notions of the real world use of percentages in any regular or systematic manner. Thus, it could be argued that two objectives of percentage teaching—the need to highlight the reference to something and the development of adaptive expertise—were essentially missing. In summary, the episodes of the Flemish project teachers present a moderate cognitive complexity supported by a moderately high didactic coherence.

6.2.2 England

English students experienced conceptual knowledge, procedural knowledge, problem-solving and reasoning in frequencies commensurate, at all levels, with project norms. Taken together, and in the light of the significantly fewer codes per episode than the project norm, these allude to episodes which are both similarly and tightly structured. In addition, the effective absence of both structural knowledge and mathematical efficiency lend weight to earlier descriptions of English mathematics as pragmatically rather than theoretically oriented (Kaiser et al. 2006), concerned with functionality (Andrews 2007b), a technique-oriented subject (Leung 1995) and, in comparison with the French and German traditions, located in content simplification (Haggarty & Pepin 2002; Jennings & Dunne 1996; Kaiser et al. 2006). Furthermore, our findings suggest a learning experience in which adaptive expertise seems an unlikely outcome not least because problem solving, when it was observed, was only rarely alongside mathematical efficiency and almost never with structural knowledge. This tends to indicate that problem solving—the use of non routine problems—almost always exploited only knowledge and skills related to the topic being taught and was not used to forge links between topics. Such outcomes resonate closely with a recent schools’ inspectors’ report on the quality of mathematics teaching in England, which found a high proportion of English lessons “characterised by the teacher doing most of the talking, emphasising rules and procedures rather than concepts or links with other parts of mathematics” (Office for Standards in Education 2008, p. 20). Moreover, “practising only one method at a time... had the effect of fragmenting the subject, because it was presented as a collection of apparently arbitrary rules that had to be memorised”.

In respect of the teaching of percentages, the qualitative analysis confirmed much of the above. For example, the teacher made strong links between fractions and percentages by encouraging students to convert *easy percentages* into fractions in their lowest form (Depaepe et al. 2005). However, this sense of conceptual linking was largely undermined by an emergent emphasis on relating percentage calculations to an initial calculation of 10% which the teacher presented procedurally, by means of demonstrations and tightly focused closed questions, as dividing by ten. In the following lesson, to her surprise, several students, when asked how they would calculate 20% of something, suggested they would divide by 20. Such approaches, reflecting a systemic encouragement of informal strategies, appear not only to encourage learner acquisition of misconceptions but also fail to acknowledge important mathematical generalities. Indeed, Depaepe et al. (2005) noted that the English students were never expected to work with the conceptually more general process of dividing by 100 to find 1% of an amount. Lastly, the tasks presented to students were routine and rarely related to anything other than simple contrived contexts.

The above reflects, in the presentation of mathematics as a largely unconnected and particularised collection of facts, rules and skills, Ernest’s (1989) instrumentalist tradition.

Moreover, in respect of the particularities of percentage teaching, conceptual simplification, no contextual reference and no encouragement of adaptive expertise collude in supporting our interpretation that the episodes of English project teachers present a low cognitive complexity and barely moderate didactic coherence.

6.2.3 Hungary

In respect of their observable learning outcomes, and despite proportions of procedural and conceptual knowledge commensurate with project norms, the evidence suggests that the higher order outcomes of reasoning, structural knowledge, problem-solving and efficiency, particularly in the regular juxtaposition implied by the second level analysis, were a significantly more frequent element of Hungarian episodes than elsewhere. These findings accord with earlier accounts in which the intended (Sutherland 2000) and implemented (Andrews 2003, 2007b; Graham, Rowland, Jennings & English 1999) mathematics curricula are “focused on ... the skills of logical reasoning, which are developed through the solution of non-routine problems that allow for multiple solution-strategies and encourage higher-level thinking” (Andrews 2007b, p. 330).

This sense of mathematical complexity was supported by the qualitative analysis of the teaching of percentages. Depaepe et al. (2005, p. 167) commented that Hungarian students were engaged at a much higher conceptual level than students elsewhere, being “strongly encouraged to apply known procedures flexibly and meaningfully to new and unfamiliar tasks”. For example, they were expected to solve problems with different conditions prevailing, as on the occasions they were asked to find the original amount having been given the outcome and percentage rate. On another occasion Hungarian students were asked to consider the effect of taking 120% of 80% of something. The class discussed the problem in a general sense before examining the particular case of 100 to confirm not only their prediction of a reduction but also that the order was immaterial. Moreover, “Hungarian students learned that percentages are always related to ‘something’ and that they have no meaning without taking into account to what they refer” (Depaepe et al. 2005, p.167) with many of the problems posed to them located in analyses of global issues concerning poverty, health and illiteracy. Thus, in terms of the learning objectives associated with percentages, it would seem that Hungarian teaching encourages both conceptual and procedural knowledge and that these are managed in ways that facilitate adaptive expertise and an understanding of percentages as being of the real rather mathematical world.

In summary, the high levels of structural knowledge and mathematical efficiency, when viewed alongside the significantly higher proportions of reasoning, tend to reflect a Platonist perspective on the subject. Moreover, the frequency of problem solving, while not entirely distinct from Ernest’s (1989) problem-solving conception of mathematics as “a dynamic, continually expanding field of human creation and invention, a cultural product” underpins a systemic view that mathematics is there to be discovered through the use of judiciously selected problems (Andrews 2003; Szalontai 2000). Thus, in conclusion, the episodes of the Hungarian teachers of this study, when considered against the higher than project norm number of codes per episode, present a high cognitive complexity supported by high levels of didactic coherence.

6.2.4 Spain

Spanish episodes attracted significantly more codes than the project norm. They also yielded stronger emphases on procedural knowledge and problem-solving than elsewhere and although the remaining outcomes were observed in proportions commensurate with

project norms their distributions were not. For example, mathematical efficiency was usually associated with conceptual knowledge, procedural knowledge or both but rarely with problem solving. This indicates limited expectations in respect of how non-routine problems are used and may reflect Blanco's (2003) observation that Spanish teachers have idiosyncratic views about the nature of mathematics and its teaching. Also, a relative lack of structural knowledge with either procedural knowledge or problem solving indicate that non-routine problems rarely, if ever, cross topic boundaries. Thus Spanish episodes appear focused on narrow applications of mathematical knowledge. However, such applications are not necessarily real world. Indeed, in none of the 21⁴ Spanish videotaped lessons did we see "teachers' consistent linking of principles studied with practical, everyday applications" as found by the survey of mathematics and science opportunities (SMSO) team (Schmidt et al. 1996, p. 99). An explanation for this discrepancy might lie in the autonomy of the Spanish regions whereby each is free to adapt centrally determined curriculum goals to local needs (Alsina & Richard 2001) and the fact that the METE lessons all took place in south-west Andalucía, a region of Spain with strong local traditions.

In respect of the percentage lessons the above finds resonance with Depaepe et al's (2005) comment that the Spanish teacher placed a "strong emphasis on computational goals" alongside low level procedures like converting "percentages into equivalent fractions in order to calculate the percentage" (Depaepe et al. 2005, p. 166). That said, students were gradually guided into calculating percentages "by multiplying the given percentage by 1% of the given amount". However, "adaptive expertise was not aimed at, since students were always asked to apply the same procedure to familiar tasks" and, while the teacher introduced the topic through a discussion focused on children's understanding of the real-world use of percentages, notions of real-world relevance were not revisited during the same or subsequent lessons. Thus, in terms of the goals of percentage teaching, both adaptive expertise and contextual reference were lacking in the lessons observed.

The unusually high levels of problem solving allude less to Ernest's (1989) problem solving conception than, in its narrow focus and juxtaposition with procedural knowledge, to his instrumentalist. This, linked to the lack of structural knowledge and mathematical efficiency suggest that mathematics is viewed as having little intrinsic value. Indeed, it could be argued that the Spanish tradition is one of application with no other objective but application. Thus, the episodes of the Spanish project teachers comprise low cognitive complexity allied to low didactic coherence.

6.3 Instrument potential

In terms of the third issue, establishing the potential of METE-like instruments for comparative mathematics education research in general and, by implication, examination of the national teaching script in particular, we feel we have been successful. The generic learning outcomes discussed above have shown, apart from their ease of use and potential for reliable coding, similarities and differences in the ways in which teachers from different countries present mathematics that resonate with the earlier studies available to us. Moreover, the significance levels reported above, and the measures taken to prevent both type I and type II errors, allow for some interpretive latitude and support a conjecture that differences identified by the analyses were unlikely to have been due to chance. In short,

⁴ The project team videotaped five lessons during the Spanish observation week and 16 during the formal round of topic-based data collection.

the seven generic learning outcomes and the emergent constructs of *cognitive complexity* and *didactic coherence* have enabled the development of tentative models of nationally located expectations of mathematics teaching.

7 Concluding thoughts

Our objective in this paper was to validate the project's instrument. In this respect, at least in terms of teachers' observable learning objectives, we believe we have been successful. Firstly, the manner of the instrument's development and the involvement of a multinational team working in the bottom-up manner described above, imparted considerable content validity. Secondly, despite relatively little in respect of Flanders and Spain, the available literature accords well with the findings yielded by project data and offers further support for the construct validity of the coding schedule. Thirdly, despite the small samples, the statistical analyses warranted the instruments as a potentially powerful tool for comparative mathematics education research. Also, the limited evidence yielded by the project data confirms an initial conjecture that even culturally near-neighbours develop significantly different educational traditions. That said; a more complete account of teachers' actions necessitates complementary research that goes beyond the sort of quantification presented here. Generic codes can tell us, in ways that allow for meaningful comparison, what is happening in a country's classrooms but offer little in respect of explanation. Therefore, attention should be paid not only to socio-cultural and politico-historical analyses of educational systems and traditions but also to studies that privilege the voice of individuals.

Acknowledgements The mathematics education traditions of Europe (METE) project team gratefully acknowledges the financial support of the European Union, Socrates Action 6.1 programme, project code 2002-5048.

The METE project team were Erik De Corte, Fien Depaepe, Peter Op't Eynde and Lieven Verschaffel, The Catholic University of Leuven, Belgium; Paul Andrews, Gillian Hatch (sadly deceased) and Judy Sayers, Cambridge, Manchester Metropolitan and Northampton Universities respectively, England; George Malaty and Tuomas Sorvali, University of Joensuu, Finland; Kati Fried, Sári Pálfalvi, Éva Szeredi and Judit Török, Eötvös Loránd University, Budapest, Hungary; and José Carrillo, Nuria Climent and Cinta Muñoz, University of Huelva, Spain.

References

- Alsina, C., & Richard, P. (2001). *Reference levels in school mathematics education in Europe: National presentation Spain*. Paper presented to the Committee on Mathematics Education, European Mathematical Society.
- An, S., Kulm, G., & Wu, J. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education*, 7(2), 145–172.
- Andrews, P. (2003). Opportunities to learn in the Budapest mathematics classroom. *International Journal of Science and Mathematics Education*, 1(2), 201–225.
- Andrews, P. (2007a). Negotiating meaning in cross-national studies of mathematics teaching: kissing frogs to find princes. *Comparative Education*, 43(4), 489–509.
- Andrews, P. (2007b). The curricular importance of mathematics: a comparison of English and Hungarian teachers' espoused beliefs. *Journal of Curriculum Studies*, 39(3), 317–338. doi:10.1080/00220270600773082.
- Baroody, A. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Mahwah: Lawrence Erlbaum.
- Blanco, L. 2003. The mathematical education of primary teachers in Spain. *International Journal of Mathematics Teaching and Learning*. Retrieved September 14, 2008, from: <http://www.cimt.plymouth.ac.uk/journal/blanco1.pdf>

- Cai, J., & Wang, T. (2006). U.S. and Chinese teachers' conceptions and constructions of representations: a case of teaching ratio concept. *International Journal of Mathematics and Science Education*, 4(1), 145–186. doi:10.1007/s10763-005-9006-7.
- Carrillo, J., & Climent, N. (2005). *The mathematics education traditions of Europe (METE) project: the teaching of polygons in primary schools*. Paper presented at the biennial conference of the European Association for Research on Learning and Instruction, Nicosia, Cyprus.
- Clarke, D. (2002). *The learner's perspective study: methodology as the enactment of a theory of practice*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Clarke, D., Emanuelsson, J., Jablonka, E., & Mok, I. (2006). *Making connections: comparing mathematics classrooms around the world*. Rotterdam: Sense.
- Clarke, D., Keitel, C., & Shimizu, Y. (2006). *Mathematics classrooms in twelve countries: the insider's perspective*. Rotterdam: Sense.
- Cogan, L. S., & Schmidt, W. H. (1999). An examination of instructional practices in six countries. In G. Kaiser, E. Luna, & I. Huntley (Eds.), *International comparisons in mathematics education* (pp. 68–85). London: Falmer.
- Depaepe, F., De Corte, E., Op't Eynde, P., & Verschaffel, L. (2005). Teaching percentages in the primary school: a four country comparative study. In L. Verschaffel, E. De Corte, G. Kanselaar, & M. Valcke (Eds.), *Powerful environments for promoting deep conceptual and strategic learning* (pp. 147–171). Leuven: Leuven University Press.
- De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction*, 13(4), 441–463. doi:10.1016/S0959-4752(02)00040-3.
- Ernest, P. (1989). The impact of beliefs on the teaching of mathematics. In P. Ernest (Ed.), *Mathematics teaching: the state of the art*. Lewes: Falmer.
- Givvin, K., Hiebert, J., Jacobs, J., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative Education Review*, 49(3), 311–343. doi:10.1086/430260.
- Graham, T., Rowlands, S., Jennings, S., & English, J. (1999). Towards whole-class interactive teaching. *Teaching Mathematics and its Applications*, 18(2), 50–60.
- Haggarty, L., & Pepin, B. (2002). An investigation of mathematics textbooks and their use in English, French and German classrooms: who gets an opportunity to learn what. *British Educational Research Journal*, 28(4), 567–590. doi:10.1080/0141192022000005832.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child Development and Education in Japan* (pp. 262–272). New York: Freeman.
- Hiebert, J., & Lefevre, P. (1986). Conceptual knowledge and procedural knowledge in mathematics: an introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics* (pp. 1–27). Hillsdale: Lawrence Erlbaum.
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, 30(2), 393–425.
- Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: results from the TIMSS 1999 video study*. Washington: National Center for Educational Statistics.
- Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, 10(3), 301–320. doi:10.1016/0147-1767(86)90015-5.
- Hufton, N., & Elliott, J. (2000). Motivation to learn: the pedagogical nexus in the Russian school: some implications for transnational research and policy borrowing. *Educational Studies*, 26(1), 115–136. doi:10.1080/03055690097772.
- Jennings, S., & Dunne, R. (1996). A critical appraisal of the National Curriculum by comparison with the French experience. *Teaching Mathematics and its Applications*, 15(2), 49–55.
- Kaiser, G., Hino, K., & Knipping, C. (2006). Proposal for a framework to analyse mathematics education traditions in eastern and western traditions. In F. Leung, K.-D. Graf, & F. Lopez-Real (Eds.), *Mathematics education in different cultural traditions: a comparative study of East Asia and the West* (pp. 319–351). New York: Springer.
- Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: helping children learn mathematics*. Washington, DC: The National Academies.
- Knipping, C. (2003). Learning from comparing: a review and reflection on qualitative oriented comparisons of teaching and learning mathematics in different countries. *ZDM. Zentralblatt für Didaktik der Mathematik*, 35(6), 282–293. ZDM. doi:10.1007/BF02656692.

- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310.
- LeTendre, G., Baker, D., Akiba, M., Goessling, B., & Wiseman, A. (2001). Teachers' work: Institutional isomorphism and cultural variation in the U.S. Germany, and Japan. *Educational Researcher*, 30(6), 3–15. doi:10.3102/0013189X030006003.
- Leung, F. (1995). The mathematics classroom in Beijing, Hong Kong and London. *Educational Studies in Mathematics*, 29(3), 297–325. doi:10.1007/BF01273909.
- Lin, F. -L. (1988). Societal differences and their influence on children's mathematical understanding. *Educational Studies in Mathematics*, 19(4), 471–497. doi:10.1007/BF00578695.
- Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah: Lawrence Erlbaum.
- Office for Standards in Education (Ofsted) (2008). *Mathematics: understanding the score*. London: Ofsted.
- Organisation for Economic Cooperation and Development (OECD) (2007). *Education at a glance: OECD indicators 2007*. Paris: OECD.
- Osborn, M. (2004). New methodologies for comparative research? Establishing 'constants' and 'contexts' in educational experience. *Oxford Review of Education*, 30(2), 265–285. doi:10.1080/0305498042000215566.
- Sayers, J., & Andrews, P.(2005). *The mathematics education traditions of Europe (METE) project: the teaching of linear equations*. Paper presented at the biennial conference of the European Association for Research on Learning and Instruction. Nicosia, Cyprus.
- Schmidt, W., Jorde, D., Cogan, L., Barrier, E., Gonzalo, I., Moser, U., et al. (1996). *Characterizing pedagogical flow: an investigation of mathematics and science teaching in six countries*. Dordrecht: Kluwer.
- Sharpe, K. (1997). The Protestant ethic and the spirit of Catholicism: ideological and institutional constraints on system change in English and French primary schooling. *Comparative Education*, 33(3), 329–348. doi:10.1080/03050069728406.
- Skemp, R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20–26.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455–488.
- Stigler, J., & Perry, M. (1990). Mathematics learning in Japanese, Chinese and American classrooms. In J. Stigler, R. Shweder, & G. Herdt (Eds.), *Cultural Psychology: essays on comparative human development* (pp. 328–353). Cambridge: Cambridge University Press.
- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS video study: Methods and findings from an exploratory research project on eighth grade mathematics instruction in Germany, Japan and the United States*. Washington: National Center for Education Statistics.
- Stigler, J., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87–100. doi:10.1207/S15326985EP3502_3.
- Sutherland, R. (2000). *A comparative study of algebra curricula: report prepared for the Qualifications and Curriculum Authority*. Bristol: University of Bristol, Graduate School of Education.
- Szalontai, T. (2000). Some facts and tendencies in Hungarian mathematics teaching. *International Journal of Mathematics Teaching and Learning*. Retrieved September 14, 2008, from: <http://www.cimt.plymouth.ac.uk/journal/tshungmt.pdf>
- Szeredi, E., & Török, J.(2005). *Teaching polygons in the secondary school: a four country comparative study*. Paper presented to the biennial conference of the European Association for Research into Learning and Instruction, Nicosia, Cyprus.
- Thompson, A. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching instructional practice. *Educational Studies in Mathematics*, 15(2), 105–127. doi:10.1007/BF00305892.
- Tweed, R., & Lehman, D. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *The American Psychologist*, 57(2), 89–99. doi:10.1037/0003-066X.57.2.89.
- Ulewicz, M., & Beatty, A. (2001). *The power of video technology in international comparative research in education*. Washington: National Academy.
- Van den Heuvel-Panhuizen, M. (1994). Improvement of (didactical) assessment by improvement of problems: an attempt with respect to percentage. *Educational Studies in Mathematics*, 27(4), 341–372.
- Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, 53(2), 113–138. doi:10.1023/A:1025516816886.
- Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction*, 4(4), 273–294. doi:10.1016/0959-4752(94)90002-7.

- Wilson, L., Andrew, C., & Sourikova, S. (2001). Shape and structure in primary mathematics lessons: a comparative study in the North-east of England and St Petersburg, Russia—some implications for the daily mathematics lesson. *British Educational Research Journal*, 27(1), 30–58. doi:[10.1080/01411920125004](https://doi.org/10.1080/01411920125004).
- Yoshida, H., Verschaffel, L., & De Corte, E. (1997). Realistic considerations in solving problematic word problems: Do Japanese and Belgian children have the same difficulties. *Learning and Instruction*, 7(4), 329–338. doi:[10.1016/S0959-4752\(97\)00007-8](https://doi.org/10.1016/S0959-4752(97)00007-8).