Scaling up strategies for change: change in formative assessment practices

Truus Dekker* and Els Feijs
University of Utrecht, the Netherlands

This article discusses some of the results of the CATCH (Classroom Assessment as a basis for Teacher Change) project. CATCH was meant to develop, apply and scale up a professional development programme designed to change teachers' instruction by helping them change their formative assessment practices. The authors focus on the analysis of three consecutive rounds of structured interviews with CATCH teachers in two US school districts. Regarding teacher changes through participation in this programme, changes in attitude toward assessment as well as in their classroom practices were found for all participants from both research sites. Some striking attitude changes were related to the so-called assessment pyramid, a crucial element of the CATCH professional development programme. The interviews also revealed which sources of support teachers thought were important to help sustain these changes in their classrooms, schools and school districts. Evidence of scaling up and dissemination of presented ideas, ideas ‘travelling’ to other curriculum areas, larger groups of teachers and other grade levels was also found. Appropriate organizational structures have enabled ‘travel’ on a district level.

Introduction

This article discusses some of the results of a larger study, the Classroom Assessment as a basis for Teacher Change (CATCH) project. The project was a collaborative effort of mathematics education researchers of the Wisconsin Centre for Educational Research at the University of Wisconsin-Madison and the Freudenthal Institute at the University of Utrecht, the Netherlands. CATCH was meant to develop, apply and scale up a professional development programme designed to bring about fundamental changes in teachers’ instruction by helping them change their formative assessment practices. Evidence of a change in teachers’ attitude towards assessment and their assessment practices was found by analysing successive rounds of structured teacher interviews from two school districts in the United States. The interviews also revealed which sources of support teachers thought were important to help sustain
these changes in their classrooms, schools and school districts. We also found evidence of how these ideas ‘travelled’, thus pertaining to scaling up the results of the project.

The ideas on which the CATCH project, intended for mathematics teachers at middle grades, was based and which served as a starting point for the design of professional development materials used in the project, find their origin in the work done at the Freudenthal Institute, University of Utrecht, the Netherlands.

Just as we found with teachers in the Netherlands, when the curriculum changed and learning goals became more process-oriented, teachers in the United States often showed limited understanding of formative assessment practices and, as a consequence, provide students with incomplete information about their progress—see the Research on Assessment Practices (RAP) project (Romberg, 1999). When teachers learn to utilize formative assessment practices in their classrooms as a consequence of appropriate professional development, there are positive effects on student learning and achievement (Black & Wiliam, 1998). The RAP project showed that when teachers retained conventional assessment techniques, especially those focused on assessing basic skills, they also paid little attention to different strategies used by their students to solve a problem, to classroom discussions and, in general, to ‘teaching and learning with understanding’, even when they used reformed curricula.

It is on the basis of these past experiences that we assumed that changing the assessment practices of teachers already using reformed curricula could play an important part to enhance ‘teaching and assessing for understanding’. By posing questions and using test items that evoke mathematical reasoning and generalizing and by asking students to choose their own mathematical tools to solve a problem, it becomes clear what students are able to do instead of which facts, standard algorithms and definitions they know. If the process is as important as the product, if students’ own strategies are appreciated, it becomes necessary for the teacher to listen carefully to what students say in class and to assess student work more closely. Teachers use the information gathered this way to guide instruction. Textbooks do not always provide good problems. So a crucial step for teachers to take is to adapt questions posed in the curriculum or to enhance their own assessment problems. Then the design of balanced assessments is being discussed, using problems at different competency levels. Assessing becomes a continuous process, an integrated part of the teaching and learning process instead of something from outside, interrupting this process but nevertheless inevitable. We feel that this results in a more ‘student centred instructional environment, more likely to improve student achievement’ (Bransford et al., 1999).

Professional development: underlying ideas

Studies on assessment by researchers at the Freudenthal Institute (de Lange, 1987; van den Heuvel-Panhuizen, 1996) place assessment as an important part of the teaching and learning process and emphasize assessing for understanding. An essential part of the ideas underpinning the assessment principles is the distinction of mathematical
competencies at three levels (de Lange, 1995). These ideas were further expanded into the so-called pyramid model (de Lange & Boertien, 1994; Verhage & de Lange, 1997; de Lange, 1999) used to design a National Option for the Third International Mathematics and Science Study (TIMSS) study (Kuiper et al., 2000).

These ideas, incorporated in the Framework for Classroom Assessment in Mathematics (de Lange, 1999) that forms the theoretical framework for the CATCH project, were also applied in the Organization for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) framework for the assessment of mathematical literacy (OECD, 1999).

In the pyramid model, shown in Figure 1, three levels of mathematical competency are discerned:

- Level 1: Reproduction, procedures, concepts and definitions;
- Level 2: Connections and integration for problem solving;
- Level 3: Mathematization, mathematical thinking and reasoning, generalization and insight.

A considerable part of a test will often consist of Level 1 questions that are given to students to determine whether the basic facts or skills taught previously are mastered. When answering questions at Level 2, students have to choose their own mathematical tools and use their own strategies. Problems on this level can usually be solved in several correct ways. The competencies on Level 3 include a critical component and reflection on the process. Students should not only be able to solve problems but also
to pose questions themselves and communicate using mathematical language. They must be able to use and criticize a mathematical model and remodel if necessary. The outcomes are translated back to the real life situation if needed.

Assuming that items at the reproduction level are often (closed) open or multiple choice questions, and that for Level 2 and even more so for Level 3 extended response open questions are needed that allow for more time, we often use the ratio 3:2:1 as an indication of items per Level 1, 2 and 3.

While passing through the professional development trajectory designed in the CATCH project, we expected teachers to (a) recognize problems at different competency levels, (b) use and design more higher level problems for their own assessments, (c) use instructionally embedded assessment, and (d) use a more varied set of assessment instruments in general.

Data collection and research methods

In order to answer the research questions posed within the CATCH project, data was gathered from two school districts, A and B. Data for this project based on test scores on compulsory standardized district tests, and used to show how students’ achievement changed during the project, was collected and analysed, but not reported in this paper.

Data also was gathered through an initial survey, three rounds of classroom observations and teacher interviews, through collecting assessment portfolios and finally through an exit survey. In this article, we look for answers to our research questions based on an analysis of the interviews. Observations and assessment portfolios have been used to analyse individual teacher development.

In order to analyse the three consecutive rounds of structured interviews, based on our expectations described in the previous paragraph we formulated a set of variables (codes) to determine changes in teachers’ attitude and classroom practice with respect to classroom assessment, the sources of support they thought were important and how ideas from the project ‘travelled’. Answers from teachers were then coded. Changes in teachers’ attitude or practice were established through indirect questioning, as shown by two questions from the second interview protocol:

(1) Describe how you assess student learning this school year. How does this differ from last year?
(2) How close do you think you are to using exemplary classroom assessment practices?

Based on the answers teachers gave on these questions, we were able to code whether there was a change in teachers’ views and attitude towards assessment (main code cta) and/or a change in teachers’ classroom practice (main code ccp) as compared to the previous interview. The third main code (ss, support from different sources) was used to answer the research question about support needed by school personnel and teachers. During the last interview many teachers talked about ways the ideas from the project ‘travelled’ to other grades and other subjects, so we added a special code ‘travel’ for the analysis of the third interview round.
The main codes were subdivided into sub-codes to enable a more detailed analysis of the nature of the change and the specific sources of support that were shown in the teachers’ answers; for example, teachers might show a better understanding of and more confidence in assessment issues in general (cta-und) or they might say they now used more levels of competency from the pyramid model in their tests (ccp-pyr). We expected teachers to mention for example, ‘release time’ (ss-time) as an important way of supporting their work, or ‘the GAP book’ (Dekker & Querelle, 2002), (ss-gap) as a source of background materials. An example of answers with sub-codes is given below. Each quotation is labelled with the initials of the pseudonyms of the teacher and the interview round it is selected from.

MP, INT2: I just guess I, myself, am gaining more confidence in the fact that my understanding is better and therefore, I feel that I am much more able to assess students adequately. (cta-und)

SB, INT2: I have tried to add a little bit more high-level questions on my exams but I need more practice on it. I really, really do, in the creation of those test questions. (ccp-pyr)

For the analysis of the interviews, the Multiple Episode Protocol Analysis (MEPA) program was used.

After a few trials to analyse the interviews using the codes, we found the inter-rater reliability was not sufficient; Cohen’s Kappa of 0.4 was too low. The codes were discussed and adapted by the researchers. Inter-judge agreement was eventually secured by having the data analysed and categorized independently by two members of the CATCH research team, who discussed attributed codes until perfect agreement was reached.

Furthermore new special codes were added for coding (non) success (codes ns and sa) and records that are important for comparison with earlier surveys/interviews. An example of a teacher statement coded sa (success in student achievement):

JG, INT3: Well, I definitely see progress. Just the requirements of higher-level math. Kids are doing more things than they ever did before.

Since in district B many changes took place during the second year of the project which, to their frustration, influenced the work of many teachers in this district, a code ‘circumstantial frustration’ (cf) was added for coding the last interview. The final version of the complete list of codes is shown in Figure 2.

Note that all codes were assigned as a comparison to the situation at the start of the project or the previous interview. The total number of interviewed teachers stayed the same though a few changes occurred due to teachers moving to other schools or districts.

Analysis of interview data: results

The program MEPA, used for the analysis of the interviews, provides the opportunity to count the frequencies in which each of the codes was attributed to episodes of the interview transcripts. This enables us to answer questions such as: how does the
number of times each of the codes was attributed change over time comparing the three rounds of interviews? For how many teachers was each of the codes found in the three consecutive interviews? What are the differences of code frequencies between the two research sites A and B? Because of the relatively small number of teachers (12) who took part in the interview rounds, we decided not to use other statistical measures but to base our conclusions only on the absolute frequencies.

Another way to describe the data is based on the qualitative analysis of actual expressions by interviewees. Since the use of MEPA enables us to produce selections
of protocol episodes on the basis of attributed codes we may summarize or give examples of things teachers actually said during interviews.

In the following description of our findings, besides the frequencies found for each code in the three interview rounds, some selected examples from interview transcripts will be given to deepen the insight. After a discussion of the findings our conclusions will be summarized.

One of our research questions was: how do teachers’ assessment practices change as a result of their participation in this professional development programme? In order to answer this question we first look at two main codes as listed in the overview of codes in Figure 2: changes in teachers’ attitude toward assessment (cta) and changes in classroom practice (ccp). Table 1 shows the frequencies in which these two codes were attributed. Not only the numbers of teachers who were ‘positively coded’ are listed, also the number of times each of the codes was found. For the fact that there are several distinct instances in which a teacher during one interview says things we considered to deserve the same code gives an indication of its importance in the case of this particular teacher. Besides, the frequencies for both codes cta and ccp are add-ups from their related sub-codes, as listed in Figure 2, so for example, ccp-asw, ccp-ie, ccp-lt, etc., all attribute to the frequency count of main code ccp.

From this table we learn that during the first interview a change in teachers’ attitude was found for all participating teachers, whereas a change in classroom practice was reported for half of them. Taking into account that the first interview round was conducted only two months after the CATCH project started with the initial professional development the fact that 6 out of 12 teachers already reported some sort of change in their classroom practice with regard to assessment may be considered remarkable. The fact that all teachers were found to have changed their attitude is equally striking.

Both following rounds of interviews show changes in attitude as well as in classroom practice for all of the participating teachers, according to the table, with one exception. Looking in detail at the code frequencies of the one teacher who did not show a change in attitude during the last interview brings us to the conclusion that she had already incorporated all of her changed ideas into her classroom practices.

Looking at the number of times each of the main codes was attributed in the three rounds of interviews, as shown in Table 1, we see a considerable increase in the time span between Interview 1 and Interview 2 for both cta and (especially) ccp, implying that teachers by that time had managed to bring changed attitudes into practice.

<table>
<thead>
<tr>
<th></th>
<th>Interview 1</th>
<th>Interview 2</th>
<th>Interview 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cta</strong></td>
<td>12 teachers</td>
<td>12 teachers</td>
<td>11 teachers</td>
</tr>
<tr>
<td></td>
<td>47 times</td>
<td>71 times</td>
<td>40 times</td>
</tr>
<tr>
<td><strong>ccp</strong></td>
<td>6 teachers</td>
<td>12 teachers</td>
<td>12 teachers</td>
</tr>
<tr>
<td></td>
<td>6 times</td>
<td>91 times</td>
<td>49 times</td>
</tr>
</tbody>
</table>

Table 1. Frequencies for codes cta and ccp: number of teachers and number of times.
Interview 3 shows that the number of times a changed attitude was found is a little below the level of the first interview and that the changing of classroom practice is still very high if compared to the first interview, but went down considerably. Searching for an explanation for this decrease in changed practices brought us to examining possible differences in the frequencies found in research sites A and B.

Table 2 shows how the numbers of times the codes cta and ccp were found (as reported in Table 1) are distributed over the two research sites A and B. We see, for example, that for the second round of interviews the number of times both codes were found is fairly evenly distributed over the two sites. In the first interview round the number of times changes in teachers’ attitude was found in Site B is comparable to that in Site A, taking into account that there were twice as many teachers in Site B than in Site A. During the time of the first interview, teachers in Site B appeared to be a little ahead as far as change in classroom practice is concerned. However, considerable differences between the two sites are found in the frequencies for the third interview round. Whereas the change in teacher attitude remains high in site B, it becomes clear that the decrease in changed classroom practices is to be attributed to the results of the interviews with teachers in site B. This decrease may be satisfactorily explained by the major organizational change that took place in Site B, which meant that all but one teacher involved in the CATCH project had been transferred to privatized schools with new management by the time the third interview took place. The only lead teacher still working in a public school had left her school (and the school district) by that time. As teachers from this site reported in the third interview, the organizational transfer came with a change in curriculum and management, not allowing for implementation of the ideas the teachers had developed ownership of during their participation in the CATCH project. We illustrate this with a few quotations coded cf (circumstantial frustration, see Figure 2).

MK, INT3: Oh, this school year. Okay. Well, this school year is a little different from other school years because we have a completely different programme. We are kind of like trained, trying to find my way around it, things I have used in the past don’t necessarily always work here. When we say, you have to assess what you teach, and what we have been teaching pretty much is basic facts, you know. Skills, things like that.

Table 2. Distribution over sites A and B of frequencies of codes cta and ccp

| Table 2. Distribution over sites A and B of frequencies of codes cta and ccp |
|---|---|---|
| Interview 1 | Interview 2 | Interview 3 |
| 4 teachers Site A | 5 teachers Site A | 5 teachers Site A |
| 8 teachers Site B | 7 teachers Site B | 7 teachers Site B |
| cta | 47 times: | 71 times: | 40 times: |
| | 15 in site A | 33 in site A | 10 in site A |
| | 32 in site B | 38 in site B | 30 in site B |
| ccp | 6 times: | 91 times: | 49 times: |
| | 1 in site A | 46 in site A | 33 in site A |
| | 5 in site B | 45 in site B | 16 in site B |
So I have been doing that, like throwing some things in that are a little bit different. But there is not much room in our curriculum for that kind of stuff. And it is kind of unfair to do it to the kids. But it is very basic. Very Level 1.

**MP, INT3:** I haven’t been able this year so far to put, to place some of that assessment that we learned all about. That is going to be a big experiment, to try to pull, I have got to develop other kinds of questions go along with this programme, because so far is it just like there is no real thinking questions, no real deep thought of questions. Nothing. There is just nothing there. So I guess I am going to try to make something there that is not there. Make some more problems for them to think and ponder.

**Interviewer:** How close do you think you are to using exemplary classroom practices?

**EW, INT3:** In honesty, I think this year, not very close at all. Because I feel compelled to, well, they tell me I have to follow their design. But I know in time I will be able to resurrect something out of the rubble.

These quotations explain how changed circumstances may hinder implementation of a changed attitude toward assessment. Note that the last sentence shows there is still hope.

The discussion of the results of the analysis of teacher interviews so far has shown that teachers’ attitudes toward classroom assessment, as well as their classroom practices with respect to assessment in general, have changed as a result of their participation in the professional development programme (CATCH). We will now look in more detail at what these changes consist of. Table 3 shows the frequencies of the sub-codes as listed and described in Figure 2. Please note that the numbers stand for the number of teachers the codes were attributed to, not the number of times each code was found.

Table 3 shows that the most striking change in teacher attitude that had occurred by the time they were interviewed for the first time was related to the assessment pyramid. Of the interviewees 9 out of 12 were attributed the code cta-pyr. This is not surprising since the assessment pyramid and the use of more levels of assessment received a great deal of attention during the initial professional development institute. During both consecutive interviews teachers kept mentioning the impression the ideas related to the pyramid had made on their thinking about classroom assessment, thus showing equally high frequencies for Interviews 2 and 3.

Three quotations are meant to illustrate this point:

**MP, INT2:** I think we probably can’t do enough repeating about the pyramid and go back to it and re-examine it, looking at it, talking about it more. And I think, like what I said before, too, was just more, even then more examples of levels of questioning. More. More examples. A lot more. Even for myself a lot more.

**MK, INT3:** The idea of questions being at different levels. Really, it is not something that was ever addressed in teacher school or college like that. And there is so, there is so much that can be found out by creating questions that are in that, the whole idea, like the mathematizing where...
T. Dekker and E. Feijs

FS, INT3: I never knew about ‘balanced assessments’ before. I never knew that there should be certain. I knew the kids needed to explain their answers more in math and do more writing of their answers. But, I never even understood the concept, you know, of this is Level 1 basic things. And then from what I taught the children or what I am teaching them, they should now be able to answer the higher question on their own, without my assistance. Now, take what you learn. And work this problem. And that is something you have to be taught. You can’t assume. Just like with the children. You can’t assume that the teachers know. And I never knew until CATCH opened my eyes to that.

As far as the implementation of ideas related to the assessment pyramid is concerned we learn from Table 3 that the number of teachers managing to change their related classroom practices was growing: starting with 3 out of 12 teachers at the time of Interview 1, rising to 8 out of 12 teachers at the time of Interview 3. That implementation of the assessment pyramid is not at all easy may be illustrated by the following quotations selected from the three interviews:

PL, INT 1: The pyramid has really been good for me. I am trying to get things at Level 2. I have incorporated that in every chapter. I am trying to get Level 3. I am starting what I thought was Level 3. I am trying to learn from that. The students are trying to get through that, learn from that. That has been probably the most useful so far.

GC, INT2: Um. I have tried to add a little bit more high-level questions on my exams but I need more practice on it. I really, really do, in the creation of those test questions.

Table 3. Frequencies for subcodes of codes cta and ccp: number of teachers

<table>
<thead>
<tr>
<th>Code</th>
<th>Interview 1</th>
<th>Interview 2</th>
<th>Interview 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>cta-asw</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>cta-ie</td>
<td>5</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>cta-lt</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>cta-oai</td>
<td>6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>cta-pyr</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>cta-sr</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>cta-und</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>ccp-asw</td>
<td></td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>ccp-ie</td>
<td>1</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>ccp-lt</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ccp-oai</td>
<td>2</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>ccp-pyr</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>ccp-sr</td>
<td></td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Change in formative assessment practices

FS, INT3: And I think that is what CATCH helped me to do more in my classroom, to take my kids to Level 2, and sometimes Level 3. Even though it is uncomfortable to be in Level 3, it is important to have a balanced assessment.

There are some other insights to be gained from Table 3. As far as changes in teacher attitude are concerned the second major change is that during the first interview half of the teachers reported they had started to critique their own assessment instruments and were willing to use other formats (cta-oai). This number of teachers rose to 8 out of 12 in the second interview and dropped to 5 out of 12 in the last interview. This drop may be explained by a rising ability of teachers to change their classroom practices accordingly.

The third major change in teacher attitude that Table 3 draws our attention to is the change in teachers’ thinking about instructional embedded assessment and the increasing importance of observations and discussion (cta-ie). This code is attributed to 5 out of 12 teachers in the first interview, growing to 9 out of 12 for the second interview. The fact that 10 out of 12 teachers are found to have also implemented instructional embedded assessment (ccp-ie) at the time of the second interview tells us that the implementation of these ideas is a smaller step for teachers to take than implementing the assessment pyramid. However, between Interviews 2 and 3 (a time lapse of approximately 9 months) the number of teachers who had managed to implement the pyramid and instructional embedded assessment had become equally high (8 out of 12 teachers). A quotation may illustrate how one teacher described her changed classroom practice with regard to instructional embedded assessment:

JB, INT3: But in the classroom itself, I walk around to the small groups. I talk with the children, find out what is going on. By working with that small group I can see what they need. If they need to move on, if they need to take a step back, I can regroup children. So I am constantly doing all sorts of assessment as the class is going on.

The fourth change in teacher attitude that is fairly high concerns teachers’ attention to the analysis of student work and strategies, scoring and grading, and record keeping (cta-asw). In the first interview 4 out of 12 teachers reported that their attitude had changed in this respect. At that time we found no signs of implementation of this grown attention, but at the time of the second interview 10 out of 12 teachers were found to have adjusted their classroom practices. One teacher described her changed practice with regard to assessment of student work as follows:

PL, INT2: I can read the kids better than I could two years ago. And to the kid instead of, you know, it used to be as general as a class it seemed okay. It was a class for not doing as well. I think I can pretty much to the kid know what they are doing.

We conclude the discussion of the data presented in Table 3 with some final remarks. We had expected teachers to change with respect to the number of tests, quizzes and homework checks (cta-lt and ccp-lt), but found only few teachers reporting that they used these instruments less frequently. Another expected change
concerned the emphasis on student responsibility (cta-sr and ccp-sr), which apparently was not high on teachers’ list. And, finally, we find that the number of teachers reporting that in general they had gained a better understanding of and more confidence in assessment issues had grown to 6 out of 12 during the last interview. As one teacher put it:

GC, INT3: The fact that I am looking at assessment very, very differently now. It really did alter me dramatically in that. I may have not made as many steps as I want to in the process of change, but my mindset is changed. And I know that over time it will be more effective.

We will now continue with another research question addressed in this article: what support do school personnel and teachers in various school contexts, who are adapting these principles to local conditions, need to ensure that changes in formative assessment are sustained?

In the analysis of the transcripts of all three interview rounds we have coded all references by teachers to any source they experienced as a support in their change processes, whether they were specifically asked about support through questions in the structured interview or whether they offered such information while talking about other issues. In total, that is for both Site A and B, we have found 67 references to sources of support during the first interview, 88 references in the second interview, and 52 references in the third interview. Since we were interested in the sort of support that was found to be helpful by teachers participating in the CATCH project the sub-codes as presented in Figure 2 were used in the coding of the interview transcripts. Table 4 shows the frequency in which each so-called source of support (ss) was found, the numbers standing for the number of teachers making a reference to a specific source.

The support participating teachers experienced from their colleagues, amongst them head teachers (ss-col), absolutely stands out according to Table 4. And if we look at the data from the two sites A and B (not given in this table) there seems to be no

<table>
<thead>
<tr>
<th>Table 4. Frequencies for sub-codes of code ss: number of teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>ss-am</td>
</tr>
<tr>
<td>ss-catch</td>
</tr>
<tr>
<td>ss-col</td>
</tr>
<tr>
<td>ss-cur</td>
</tr>
<tr>
<td>ss-gap</td>
</tr>
<tr>
<td>ss-mat</td>
</tr>
<tr>
<td>ss-prin</td>
</tr>
<tr>
<td>ss-time</td>
</tr>
<tr>
<td>ss-utr</td>
</tr>
<tr>
<td>ss-web</td>
</tr>
<tr>
<td>ss-work</td>
</tr>
</tbody>
</table>
difference in this respect. From the interview transcripts we learn that being able to share with colleagues was considered to be very important, whether through professional monthly or weekly meetings in different settings or whether through frequent informal conversations. We illustrate this with one quotation from each interview round:

JB, INT1: Talking with different teachers. What works good for them may also help me or maybe what I am using might help them. Just networking with math teachers throughout this cluster and throughout the district.

MP, INT2: I think the more you get together and talk about it the more it keeps you, it certainly keeps you involved in what your goals are supposed to be with this.

LM, INT3: And the fact that we have the leadership team, the human resources. To me that is the most valuable.

Looking at all three of the interviews references to the support teachers received from the CATCH team (ss-catch) were made by a little less than half of the participating teachers. What teachers valued most, according to their remarks in the interviews, was the contact in person with members from the CATCH team, especially with the site coordinators who regularly visited each of the participating teachers, not only to interview them but also to observe their classes, attend team meetings and give input for local summer institutes. We illustrate this with two quotations, one from each research site:

MK, INT2: When you guys come in it is just, because this is a discussion that goes round and round all of the time. And that is really, that is great support, because you know, you are thinking, just give me a book where all the problems are like 3 times 4. (Laughter) But, you know, like when you have other people to talk it over with it just makes it so much, you kind of like recommit, you understand why you are doing it and you go back on course.

HP, INT3: But the CATCH, the team approach of having a person to talk with about what is happening at this time, is so valuable rather than waiting until the end of the year. And you maybe have made mistakes that you wish you would have discovered sooner or other ways of thinking about things or other methods of teaching or, so having that immediate feedback is very helpful. Talking with where my students need to be at the end of the year is extremely important because that has always sort of been a dead end.

As far as human resources are concerned, besides colleagues and CATCH team members participating teachers also considered themselves lucky to have supportive principals and administrators, or to be backed up by the director of instruction or the school board (ss-prin in Table 4).

Initially the Utrecht seminar and the information participants received about the assessment pyramid during this seminar (ss-utr) scored highly: this source of support was mentioned by 8 out of 12 teachers. Even two years after the start of the project 4 out of 12 teachers kept referring to the influence of this seminar on their thinking and class work.

The resource category workshops, summer institutes and conferences (ss-work) scored fairly highly as well. In the interviews teachers were referring to a variety of
workshops and conferences they attended, but it is noteworthy that during the third interview 4 out of 5 teachers from site A stated how important their own local CATCH summer institute had been to them. We quote one of these teachers:

**KE, INT3:** I think the in-service that we ran this summer as a teacher in the CATCH Programme. I think that was very important and helped because it gave me a chance to reflect on some of the questions and statements that were made by the teachers who were involved in that. I think in thinking through the process of how were we going to present this to them. Something that we have worked on for a year and a half and have been involved in all kinds of aspects of it, and trying to put it into a two-day conference with other teachers who have not really been involved at all. I think what it did was it gave me a chance to really evaluate what I have learned from the CATCH Programme, the involvement with working with the people from the Netherlands and from the UW-Madison. I think it has increased my knowledge. It has increased my ability to work with people.

Let us now focus on the material sources of support as they are listed in Table 4. Overall, about half of the participating teachers mentioned the importance of having curriculum materials that are in line with the assessment ideas presented in CATCH at hand to lean on. We quote one teacher:

**BC, INT3:** That has been a comfort zone for me to have a starting point that is a pretty thoughtful group of different types of assessments while I am trying to get the rest of the ducks in a row.

Overall about one third of the interviewees mentioned different types of written materials they managed to lay their hands on which they considered to be of help (ss-mat). Some materials were provided by the CATCH team, such as the GAP book that was in draft form published on the web (ss-gap) which one of the teachers referred to as a ‘refresher course’.

AssessMath! (ss-am) appeared on the list only after its introduction during a local summer institute in Site B, before the second interview was conducted. During the last interview 5 out of 7 teachers in Site B reported they found this programme very helpful as a resource for designing their own assessments.

According to Table 4 the CATCH web site (ss-web) has not had the impact we had been hoping for. Analysis of the interview transcripts reveals that this is due to the fact that the online discussions that participants had been hoping for were not realized and to participants losing interest since there were considerable time lapses before the publishing of new information on the web.

It is remarkable that release time (ss-time) was hardly considered important as a support in teachers’ change process.

We conclude our discussion of the results of the analysis of the interview transcripts by examining the outcomes of the code ‘travel’ that was added for the coding of the last round of interviews. We consider this code interesting since it pertains to the scaling up goal of the CATCH project. Remarks indicating that CATCH ideas are indeed travelling were found 10 times in the third interview round. From Site A 4 out of 12 teachers gave evidence of travelling, from Site B this was 6 out of 12 teachers.
Analysis of the transcripts indicates that there were two different types of travelling: one was the transfer taking place to other curriculum areas, the other can be called ‘scaling up’, since the transfer was to larger groups of people or to other grade levels within the mathematics area. We illustrate each of these two types of travelling with two quotations:

MK, INT3: The balanced assessment idea is like primary. It is a kind of thing we, you get all levels, all kinds of thinking, and in all, like, you know, there is that Level 1, Level 2, and the different kinds of levels, of Level 1. And all that. So there is like so much that can be pulled. And that is kind of like a model, that it kind of has guided last year. And actually that is kind of like even leaked into other curricular areas. I do that with my reading.

CG, INT3: But you know, what is really cool about this whole assessment thing? It can work in science too.

EW, INT3: Well, I share my ideas with the other teachers that teach math here as well as when I teach math education courses at XX and XX. I always use the CATCH philosophy for assessment.

KE, INT3: And so I think the change is coming. The change is here. And I think more and more people, you know, thought it was going to be isolated, middle school. It has moved in both directions. And I think more and more people are starting to pick up on it. So it is a very positive movement forward.

It has been rather unexpected, even to the members of the research team of the project, both in the Netherlands and the US, how fast ideas of the CATCH project ‘travelled’. In the previous paragraph we showed how teachers talked about the ideas from the project with their colleagues from other grades and other subjects such as reading and science. However, the interviews did not show exactly how ideas ‘travelled’ on a district level. Participants in the CATCH project contributed to both. Due to the existence of an appropriate organizational structure at the district level, plans to disseminate ideas from the project could be carried out. During consecutive local summer institutes, organized by teachers from the CATCH leadership teams, a large number of teachers from other grades and other schools took part in the professional development programme. The lead teachers also contributed to a number of conferences and seminars.

In district A, members of the CATCH team to a great extent assumed responsibility in school and district leadership by fostering consistency in classroom assessment practices at each grade level, including elementary and high school. CATCH lead teachers in this district were included in the writing of the district mathematics framework and the design of open-ended items used on district-wide assessments. In district B, ‘travelling’ on a district level has taken place to an even greater extent in the sense that a larger number of teachers became involved in the CATCH professional development programme. Other districts have shown interest in the CATCH professional development programme. The materials developed for this programme are available on the CATCH website (www.fi.uu.nl/catch).
Conclusions

Based on the analysis of the transcripts of three rounds of interviews conducted with participating lead teachers during two-and-a-half years of the CATCH project we may draw conclusions pertaining to research questions of this project.

With respect to changes in teachers’ assessment practices as a result of their participation in this professional development programme for all the participating teachers from both research sites, A and B, we found changes in their attitude towards assessment as well as in their classroom practices. A change in attitude was accomplished for all teachers shortly after the initial professional development programme was held; a change in classroom practice appears to take more time but was nevertheless reported by all participants within a year’s time. The level of change was sustained to a great extent throughout the remaining project time, with the exception of teachers in Site B who suffered from rather dramatic changes in their teaching circumstances. The most striking changes in teachers’ attitude towards assessment were related to the assessment pyramid, which was a crucial element of the CATCH professional development programme designed to support changes in teachers’ formative assessment practices. Our findings indicate that ideas related to instructional embedded assessment and the critiquing of teachers’ own assessment instruments and trying other assessment formats were more easily implemented, but nevertheless over time there was a growing number of teachers managing implementation of assessment pyramid related ideas.

Analysis of the interviews offers detailed information about the support teachers need to ensure that changes in formative assessment practices are sustained. The most outstanding source of support we found was the frequent personal contact with colleagues, whether through professional meetings or through informal contacts. Participants also highly valued the information received at the initial professional development seminar in Utrecht, and the contact with members of the CATCH team in person through regular site visits which to teachers’ delight included opportunities to attend and discuss their classes. Furthermore, in the interviews teachers mentioned the support of different types of materials they received and indicated how important the organization of their own local summer institutes had been in their professional development.

Finally, with respect to dissemination of professional development ideas and scaling up we found evidence of ideas ‘travelling’ to other curriculum areas as well as to larger groups of teachers and other grade levels. Appropriate organizational structures have also enabled travel on a district level.

Notes

2. Research on Assessment Practices was one of the projects of the National Centre for Improving Student Learning and Achievement in Mathematics and Science (NCISLA), funded by OERI,
3. District A is a small urban/suburban district, serving over 3,000 students predominantly European American (85%) with approximately 30% free or reduced-cost meals. District B is a large urban district in Eastern United States, where middle grade teachers work with predominantly African American and Hispanic students, with more than 75% receiving free or reduced-cost meals.

4. In general, student achievement scores on standardized tests improved and students taught by CATCH teachers had higher scores than those taught by non-CATCH teachers.

5. Observation protocols were adapted from instruments used by Horizon Research, see http://www.horizon-research.com/LSC/manual/0203/existing.php#6

6. The observations do not contradict the analysis of the interviews, they seem to support its results.


8. MEPA, designed by Gijsbert Erkens, University of Utrecht, the Netherlands, is a flexible program for the coding of verbal data, in our case being the transcripts of the conducted interviews. The program can be used to analyse the coded data with a variety of quantitative and qualitative methods.

Notes on contributors

Truus Dekker is a senior staff member of the Freudenthal Institute, University of Utrecht, the Netherlands, and has over 25 years’ experience as a secondary mathematics teacher. Her recent work at FI focused on assessment development. She is currently involved in writing materials for an NSF funded mathematics middle school curriculum. Truus has been a speaker at many national and international assessment workshops.

Els Feijs is a researcher at the Freudenthal Institute, University of Utrecht, the Netherlands. Over the last twenty years she has been involved in various activities at this institute as a researcher, educator and curriculum developer, including professional development for both teacher trainers and in-service mathematics teachers and the development of an NSF funded mathematics middle school curriculum.

References

de Lange, J. (1999) Framework for classroom assessment in mathematics (Utrecht, Freudenthal Institute & University of Madison, WI, National Centre for Improving Student Learning and Achievement in Mathematics and Science).


