

Video technology in the assessment of an in-service teacher learning program –

Differences in mathematics teachers' judgements on instructional quality¹

Sebastian Kuntze, University of Munich (Germany)

***Abstract:** Video technology offers the possibility to make instructional situations available for discussion in teacher learning projects. For the confrontation with videotaped instructional situations, the teachers' rating of characteristics for instructional quality plays an imminent role. As criteria for instructional quality are often linked to the goals of the teacher learning project, the teachers' views offer possibilities to evaluate these projects. For the example of instructional situations in German classrooms concerning geometrical proof, differences in judgements on instructional quality are analysed. The study focuses on data of a cluster analysis showing initial divergencies in the rating of videotaped instructional situations and it describes how the teachers' views evolve.*

ZDM-Classification: B59, C79

1. Theoretical background

1.1 Videotaped instructional situations in in-service teacher learning projects

Video technology offers the possibility to show instructional situations in a relatively detailed format as far as the situational context is concerned (cf. Petko, Waldis, Pauli, & Reusser, 2003). In comparison to transcripts, for instance, videotaped classroom situations encompass a variety of nonverbal components of interaction in the form of visual and acoustical information. Probably for this reason, it is relatively easy to initiate instruction-related discussions in teacher learning projects on the base of videotaped instructional situations. One of the advantages specific for this medium is that videotaped situations may be viewed repeatedly, so that interest-

ing interactions can be examined very closely (cf. Beck, King, & Marshall, 2002).

When teachers are confronted with videotaped instructional situations and make sense of them, components of professional knowledge are involved in a bi-directional way: On the one hand, the process of making sense of videotaped situations is supposed to activate situation-specific, content-specific and more general instruction-related cognitions and individual convictions (cf. Kuntze & Reiss, 2005; Kuntze, submitted). For example, it is very probable, that an implicit or explicit judgement on the quality of instruction of the videotaped classroom situation is constructed by the teachers. The activated components of professional knowledge might influence these judgements on instructional quality.

On the other hand, the confrontation with videotaped classroom situations and their discussion can influence instruction-related cognitions and beliefs. For example, individual knowledge linked to instructional practice could be enlarged. This is one of the reasons, why changes in the domain of professional knowledge are the goal of a number of video-based in-service teacher learning projects.

As components of professional teacher knowledge very probably interfere in the video-based work of in-service teacher learning projects, it is necessary to take a look on theoretical and empirical findings concerning different domains of professional knowledge (Shulman, 1986). As professional knowledge seems to be organised episodically (Bromme, 1992, 1997; Leinhardt & Greeno, 1986), it seems necessary to consider situation-specific instruction-related cognitions and beliefs together with more general constructs (e.g. cf. Staub & Stern, 2002; Stern & Staub, 2000; Peterson et al., 1989; Lipowsky et al., 2003) and individual criteria for instructional quality (Kuntze, 2004b). For the research on how teachers judge on videotaped instructional situations, we developed the model shown in Figure 1 (Kuntze & Reiss, 2005; Kuntze, submitted). For ratings of videotaped instructional situations, situation-specific and more generalised components of professional knowledge can play a role.

¹ This research project was funded by the Robert-Bosch-Stiftung (Bew.-Nr. 32.5.8050.0037.0).

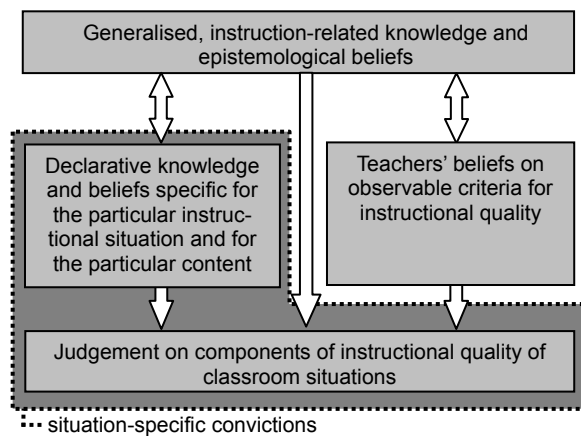


Fig. 1: Model for possible interactions between judgements on classroom situations and components of professional knowledge.

1.2 Evaluational research on a video-based teacher learning program

On the background of this model, the aim of the evaluational research on a video-based in-service teacher learning project was to find out about components of professional knowledge before the beginning of the video-based work, and to measure possible changes in general and situation-specific domains of instruction-related convictions after the teacher training. As aspects of instructional quality played an important role in the “treatment” of the teacher learning project, the teachers’ judgements on instructional quality of videotaped classroom situations were of primordial interest for evaluating the teacher training. The theoretical background of the investigation of the teachers’ criteria for instructional quality and also for the treatment in the in-service teacher learning project are empirical findings of Clausen, Reusser & Klieme (2003, cf. Klieme, 2002), who identified four “basic dimensions” of instructional quality using high-inference ratings of the sample of the TIMSS 1995 Video-Study.

In the in-service teacher learning project, we focused on instructional situations dealing with geometrical proof, in order to profit from empirical findings linked to that content (Reiss, 2002; Reiss, Hellmich & Thomas, 2002; Kuntze & Reiss, 2004; Kuntze, Rechner & Reiss, 2004). For the ratings of instructional quality in the videotaped classroom situations about geometrical proof, we concentrated on criteria in the focus domains of “cognitive activation of the stu-

dents” (cf. Clausen, 2002), “intensity of argumentation”, and “learning from mistakes” (cf. Heinze, 2004).

These characteristics of instructional quality were also in the focus of the “treatment” of the in-service teacher learning program. During three teacher training weekends, which were framing almost one school year, the teachers took part in video-based discussions concentrating on the three focus domains of instructional quality presented above. The classroom interaction was analysed cooperatively with respect to alternative teacher actions and reactions as well as their possible consequences for learning opportunities in the classroom. In the phases between the teacher training weekends, the teachers were encouraged to observe their own instructional practice and to experiment with changes according to the three focus domains (cf. Kuntze, 2004a). An improved use of the criteria of cognitive activation, intensity of argumentation and cognitively activating learning from mistakes when judging on instructional quality and the “rooting” of these criteria in the own instructional practice were central aims of the video-based teacher training.

For discussing alternatives in the instructional practice of videoed classroom situations as well as in the own classroom, the perception of instructional quality is regarded as being an important factor. Consequently, in order to evaluate the teacher learning program, it is interesting to verify whether and to what extent the judgements of the participants converge in an intended way.

According to the results of our video studies on argumentation and geometrical proof (Kuntze & Reiss, 2004; Kuntze, Rechner & Reiss, 2004) and the corresponding theoretical background (cf. Reiss, 2002), we distinguished between an instructional style marked by argumentational exchange, discourse in the classroom and cognitively activating reactions to mistakes on the one hand and a teacher-centred interaction with questions on a rather low level of complexity comparable to the dominant teacher script in Germany, as it was described in the TIMS Study (Baumert et al., 1997) on the other hand.

Corresponding to these instructional styles, two videotaped instructional situations were chosen to be rated by the participating teachers. We expected that the teachers’ judgements on instructional quality of the videotaped classroom situa-

tion marked by argumentational exchange would improve as a consequence of the treatment in the in-service teacher learning program.

For the possible relationships between professional knowledge and instructional practice, general evidence is still lacking (Tillema, 2000; Lipowsky, 2004). As a possible interpretation of the results Lipowsky (2004) reports in his overview on research on teacher learning projects, it appears that changes in professional knowledge play a role of necessary but not sufficient condition for changes in the instructional practice of teachers. Seen from this point of view, the teachers' comparisons between videotaped instructional situations and their own instructional practice might be an interesting additional indicator. For comparisons of the videotaped classroom situations to the own instructional practice we expected according to the results of the TIMS Study (Baumert et al., 1997) that before the teacher training project, the video example following the dominant German script would be perceived as rather close to the own classroom practice.

2. Research questions

The study aims at providing evidence for the following research questions:

- (i) How do teachers judge characteristics of instructional quality of videotaped instructional situations?
- (ii) How do different judgements of the teachers evolve during the teacher learning project?

3. Design of the study

Before, during and after the teacher training, 53 German and Swiss secondary mathematics teachers were asked to complete paper-and-pencil questionnaires. In order to be culture-fair, this study concentrates only on data of the 43 German participants.

Before the start of the project, a first questionnaire focused on professional knowledge and instruction-related beliefs in the sense of more general constructs (Fig. 2). For example, the teachers were asked about their cognitive constructivist or direct-transmission views on teaching and learning (Staub & Stern, 2002).

Before the presentation of the videotaped classroom situations, the teachers had to answer a second questionnaire focusing on components of their professional knowledge about introductory lessons on proof in geometry. For example, there were proof-related scales about "argumentational discourse in the classroom", "advancing by small steps" and "initial tolerance with respect to exactness" (for more detailed information, including sample items, cf. Kuntze & Reiss, 2005). Having activated their content-specific and situation-specific pedagogical knowledge by answering this questionnaire, the teachers were shown two videotaped classroom situations, both dealing with introductions to geometrical proof. According to our approach (Kuntze & Reiss, 2004; Kuntze, 2003), video A showed patterns of interaction marked by discourse and argumentational exchange between the students and the teacher, whereas video B could be characterised as a teacher-centred interaction comparable to the dominant teacher script in Germany described in the TIMS Study (Baumert et al., 1997).

Before the start of the project:

Questionnaire parallelised with the study of Lipowsky et al. (2003):
Generalised components of professional knowledge

Before the presentation of the videotaped classroom situations:

Questionnaire on situation- and content-specific cognitions and beliefs on introducing proof in geometry

Presentation of two videos showing instructional situations linked to this topic:

Video A: Patterns of interaction marked by discourse and argumentational exchange
Video B: Teacher-centred interaction corresponding to the dominant German script

Questionnaire: Judgements on instructional quality (Pretest)

After the video-based work in the in-service teacher learning project (several months later):

Presentation of the two videos showing instructional situations linked to geometrical argumentation and proof (Video A / Video B)

Questionnaire: Judgements on instructional quality (Posttest)

Fig. 2: Design of the study.

Immediately after having seen the videos, the teachers had to give judgements about these two classroom situations in a third questionnaire. In multiple-choice and open items, the teachers were asked about the particular components of instructional quality described in section 1.2, about how similar their own instructional prac-

tice was to the classroom situations and about further observations. The reliability values of the scales concerning instructional quality are shown in Table 1.

Scale	Sample Item	Number of items	Video A: Cronbach's α	Video B: Cronbach's α
cognitive activation	„The students were encouraged to learn intensively.“	3	.87	.77
intensity of argumentation	„The classroom interaction was characterised by an argumentational interchange between the students and the teacher.“	3	.84	.87
opportunities for learning from mistakes	„By the manner, in which mistakes were treated in the classroom, the students were supported to build up meaningful knowledge, that is relevant for tasks and problem solving.“	2	.80	.90

Tab. 1: Reliability values of the scales on video A and B.

After viewing the videos again several months later, which was short before the end of the project, the participating teachers were asked to answer this questionnaire another time in an enlarged version containing all basic dimensions of instructional quality. In the following, we concentrate mainly on the results of the second and third questionnaire as well as the corresponding part of the questionnaire short before the end of the project. This conforms to the interest of this study to investigate the initial content- and situation-specific convictions about instructional quality including judgements on videoed classroom situations and to make a first comparison to the teachers' ratings of instructional quality at the end of the project. For this purpose, we will first focus on differences in content- and situation-specific convictions including judgements on instructional quality of the two videos described above and gain a rough overview on the spectrum of answers using a cluster analysis.

4. Results

In order to get an overview on predispositions in the domain of instruction-related convictions of the participating teachers, a first approach was made by a cluster analysis on the base of the cognitive-constructivist or direct-transmission views of teaching and learning (Staub & Stern, 2002), of content-specific convictions about

geometrical proof in introductory lessons (Kuntze & Reiss, 2005, cf. Tab. 1) and of judgements on instructional quality in the video-taped classroom situations shown in the pretest. A two-cluster solution (Ward method) leads to the results shown in Figure 3, with a smaller group of teachers favouring video A and a larger group favouring video B, as far as the perception of instructional quality is concerned. On the left hand side, Figure 3 shows the mean values of the four scales of content-specific convictions about geometrical proof (described in more detail in Kuntze & Reiss, 2005). On the right hand side, the mean scale values for the ratings of instructional quality of the two videoed instructional situations are displayed. The means for the cognitive-constructivist or direct-transmission views of teaching and learning (Staub & Stern, 2002) are not displayed in Figure 3, because the corresponding scales range from 1 to 4. The first cluster's values were slightly more in favour of the cognitive-constructivist viewpoint (3.44 vs. 3.12 for the other cluster) and slightly lower for the direct-transmission view (2.18 vs. 2.35 for the other cluster).

While there are only tendencies of differences in instruction-related convictions between the two groups of the cluster analysis, Figure 3 shows remarkable differences in the judgements on instructional quality of the videotaped classroom situations like cognitive activation and intensity of argumentation. For this reason, we called the clusters the "group of teachers favouring video B" and the "group of teachers favouring video A".

The development of the average judgements on instructional quality from pretest to posttest is described in Figure 4. In The data in Figure 4 indicate that the different views of the two groups in the pretest might have almost vanished in the posttest. But this result should be interpreted with care: as the cluster analysis was based only on the results of the pretest, this could be due to a regression effect. For this reason, and as an additional controlling approach to the evidence in Figure 4, an additional cluster analysis was calculated focusing only on the judgements on instructional quality of the two videotaped classroom situations before and at the end of the project. Figure 5 contains the average values of the judgements in pre- and post

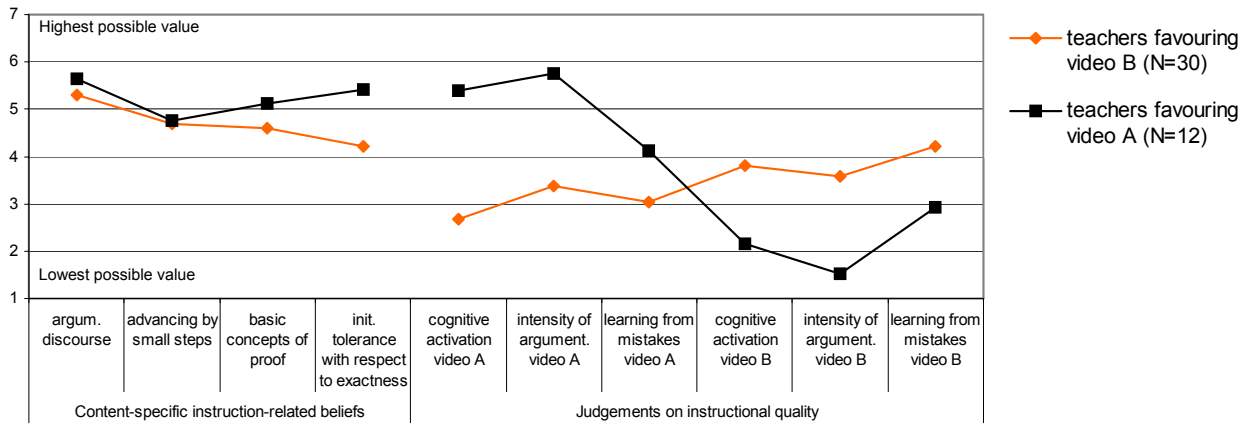


Fig. 3: Situation-specific beliefs and judgements on instructional quality (pretest).

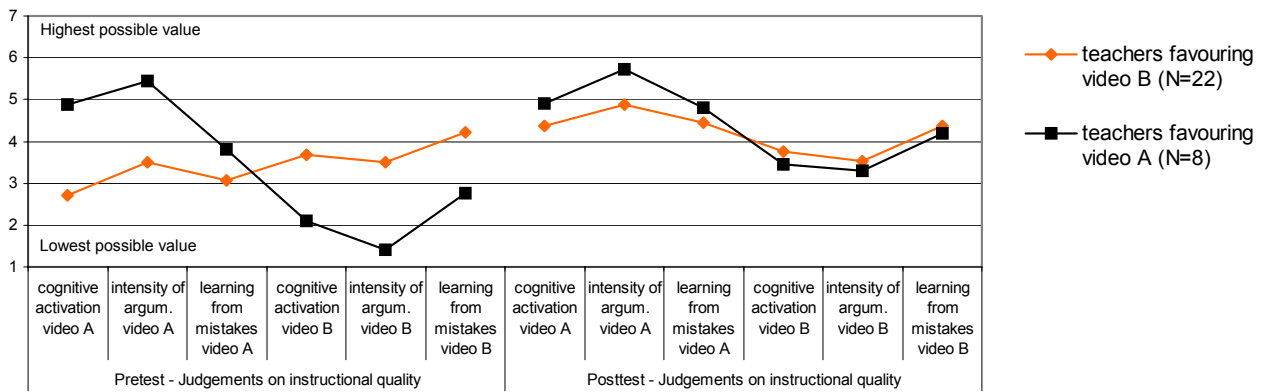


Fig. 4: Judgements on instructional quality (pretest and posttest).

test on the basis of this cluster analysis (Ward method) departing from judgement values of instructional quality in both pre- and posttest. The two resulting clusters may be described on the basis of their judgements on the videotaped classroom situations as a group of teachers with “traditionally oriented” beliefs according to their

appreciation of the dominant German teacher script (video B) before the project and a group favouring discourse as shown in video A.

In Figure 5, the basic tendency that the ratings in the pretest differed more than in the posttest is still there: While most of the judgements for the videos in the pretest differ significantly (video

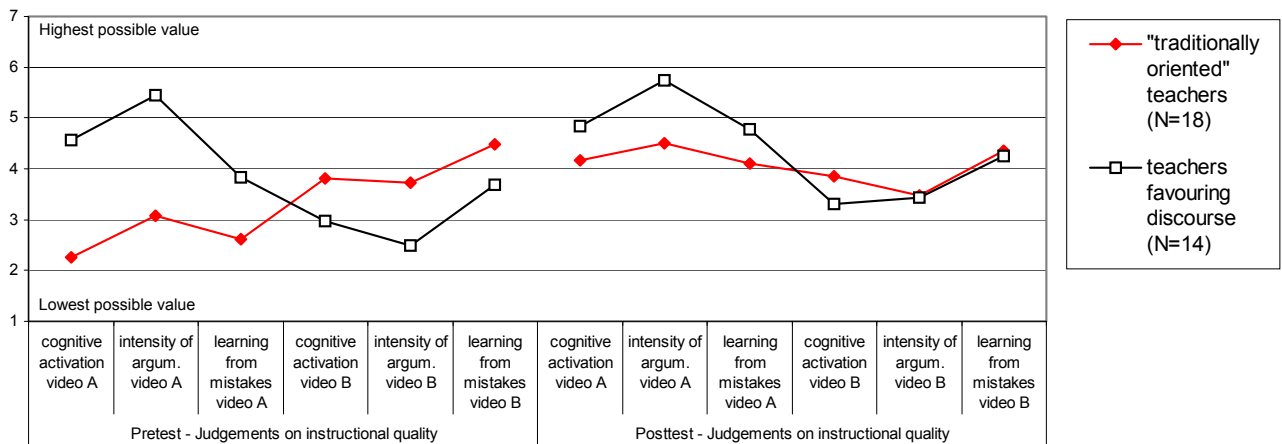


Fig. 5: Judgements on instructional quality (cluster analysis on the base of pretest and posttest).

A: cognitive activation $p < 0.001$; intensity of argumentation $p < 0.001$; learning from mistakes $p < 0.05$; video B: intensity of argumentation $p < 0.05$), there is only a significant difference for the intensity of argumentation ($p < 0.001$) in the posttest.

From pretest to posttest, we observe for video A a significantly higher rating for cognitive activation ($p < 0.05$) for the group of “traditionally oriented” teachers and more positive judgements concerning learning opportunities from mistakes for both groups ($p < 0.05$ resp.). The tendencies of “intensity of argumentation” indicate that especially the “traditionally oriented” teachers changed their views about video A.

In an additional item linked to the handling of mistakes in the classroom, the teachers were asked whether they judged situations dealing with mistakes as uncomfortable for the student who made the mistake. As shown in Figure 6,

“traditionally oriented teachers judged the situations in video B to be significantly less uncomfortable for the students having made mistakes, than did their colleagues favouring discourse ($p < 0.05$). In the posttest, there were no significant differences for this item.

Asked to compare the videotaped instructional situations to their own instructional practice (Fig. 7), both groups of teachers stated before the training, that their own lessons were more similar to video B than to video A, as well for communication and interaction in the classroom, as for the contents and tasks they would have chosen. There are tendencies that teachers favouring discourse on average see video A slightly closer to their instructional practice than the teachers of the other group do. For video B, there is an inverse tendency, with a significant difference for “contents and tasks” in video B between the clusters ($p < 0.05$). The posttest data in comparison to the pretest shows for the “traditionally oriented” teachers a highly significant change of the perceived similarity of video A for “communication and interaction” ($p < 0.01$), a significant change for “content and tasks” in video A ($p < 0.05$), and significantly less perceived similarity to “content and tasks” in video B ($p < 0.05$). The group of the teachers favouring discourse attain on average a positive value for the similarity rating in the domain of “communication and interaction” in video A, which differs significantly from the pretest ($p < 0.01$).

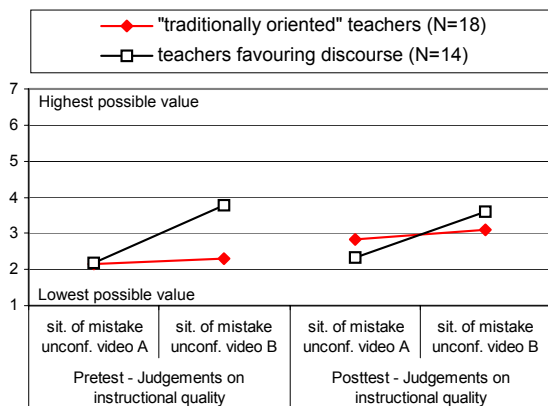


Fig. 6: Judgements on instructional quality: uncomfortable situations dealing with mistakes (pre- and posttest).

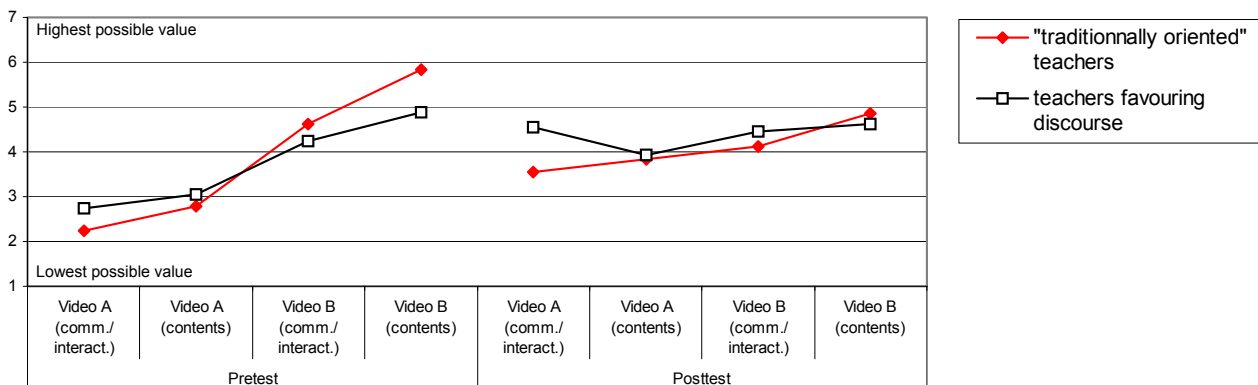


Fig. 7: Reported similarity to own instructional practice.

5. Interpretation of the results and implications for the theoretical and practical context

A first remark concerning the discussion of the results presented above is that they should be interpreted carefully. As there was no control group and the recruitment of the sample might have supported a certain selection of teachers e.g. being open towards learning from video-based teacher trainings, a generalisation of the findings seems difficult. But even if the sample of this study was not very large, some aspects of the results are interesting for the particular case of this teacher learning project and also beyond it:

The results of the cluster analyses indicate that, on the level of predispositions of the participating teachers, there were initially contrary situation-specific convictions concerning instructional quality of the videotaped classroom situations. Two conclusions may be drawn from this finding: The first one is that when confronting teachers with videotaped classroom situations in in-service teacher learning projects, we must take into account, that there might be differing judgements on instructional quality among the participating teachers. The second conclusion is that the chosen videotaped classroom situations had the potential to separate different situation-specific viewpoints, which seems a good starting point for teachers to enter a discussion about their convictions in order to elaborate shared professional knowledge on instructional quality – the differences in these judgements were much bigger than the divergences in the content-specific scales on instruction-related convictions. For the video-based work in the teacher training, we could use this controversial situation as an opportunity to elaborate criteria of instructional quality in the discussion.

The situation before the teacher training project reflects on the one hand the problem, that teachers make sense of presented videotaped instructional situations individually and differently, and on the other hand the potential of video technology to encourage discussions on instructional situations which are linked to individual convictions of the participants. For teacher learning projects, these findings highlight the importance of finding a common way with the participating

teachers of how to look at videotaped classroom situations.

The situation in the posttest seems different from the one in the pretest, and there are two main observations: The first one is that the differences in the judgements on instructional quality have been reduced remarkably. This seems to reflect better shared criteria for instructional quality, which was one of the aims of the teacher training.

The second observation concerns the changes of the views of the two groups of teachers identified in the cluster analysis. The group of “traditionally oriented” teachers, who preferred in the pretest the instructional situation corresponding to the typical German script, showed more positive ratings of the instructional quality for video A. As one of the goals of the teacher learning project was to encourage teachers to raise the intensity of argumentation, to introduce more cognitively activating tasks and to help the students to use mistakes as learning opportunities, the improved judgements of video A indicate that situation-specific components of professional knowledge might have been subject of changes as they were intended in the in-service teacher learning project.

The group of teachers favouring discourse significantly improved their ratings for “learning from mistakes”. As these teachers have reproduced their moderately positive judgements on cognitive activation and intensity of argumentation in video A, the domain “learning from mistakes” could have represented a special facet of instructional quality and improvement of instructional practice the group of teachers favouring discourse concentrated upon during the teacher learning project.

The comparisons between the videotaped instructional situations and the own instructional practice in the pretest show a higher perceived similarity to video B than to video A, like it was expected on the basis of the results of the TIMSS Video Study (Baumert et al., 1997): Video B, showing characteristics of the dominant teacher script in Germany was seen as being closer to the own instructional practice by both groups of teachers. Some tendencies in the pretest indicate that the teachers favouring discourse perceived their teaching not as close to video B, as did their “traditionally oriented” colleagues.

As a consequence of significant changes between pretest and posttest, both of the videos were rated to be almost equally similar to the own instructional practice in the posttest. Although there were plausible tendencies for differences, the two groups mainly agreed in the posttest. If we interpret these results as an indicator for the teaching of the participating teachers, we might deduce that the better ratings of the instructional quality of video A in the posttest could correspond to changes in the instructional practice: In this view, the teachers might have reported indirectly that they had approached their instructional practice to what the teachers had seen in video A and in other corresponding instructional situations which had been discussed throughout the project.

The results presented in this article lead to the question, how the initial differences in the judgements on instructional quality may be explained. First evidence for this question, indicating that the cognitive constructivist and direct-transmission views of teaching and learning could help to predict these situation-specific ratings, is given in Kuntze & Reiss (2005) and Kuntze (submitted).

Additional qualitative evidence could emerge from an investigation on the further observations of the participants concerning the videotaped instructional situations that are documented in the answers to open items contained in the third questionnaire and its posttest. A study on these items can provide qualitative insight, how the ratings of instructional quality interdependent with other professional knowledge and perceptions.

One such possible interdependence making a link between a bottom-up-analysis of individual criteria for instructional quality and the quantitative data of the third questionnaire could contribute to an explanation of the different ratings of instructional quality. First signs of such links are reported in Kuntze (submitted).

It would also be interesting to compare the teachers' judgements on instructional quality to high inference ratings made by experts from a scientific point of view. Such an investigation could give additional information on which judgements are closer to a science-based position. This can also help to evaluate the teacher learning program.

Last but not least, a further question concerns the instructional practice of the teachers. Additional evidence whether the instructional practice of the participating teachers changed in the view of students or external observers would be very interesting. In the margin of the evaluational research on the video-based teacher learning program presented here, further comparisons with data gained in other questionnaires, e.g. on reported characteristics of implementation, could provide further insight.

References

- Baumert, J., Lehmann, R., et al. (1997). *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich*. Opladen: Leske + Budrich.
- Beck, R., King, A., & Marshall, S. (2002). Effect of videocase construction on preservice teachers' observations of teaching. *The journal of experimental education*, 70(4), 345-361.
- Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens*. Bern: Hans Huber.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. Weinert (Ed.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (pp. 177-212). Göttingen: Hogrefe.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen. *Unterrichtswissenschaft*, 31(2), 122-141.
- Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I – Theoretische Grundlegung, Methodik und Ergebnisse einer Videostudie. *Journal für Mathematikdidaktik*, 25(3/4), 221-244.
- Klieme, E. (2002). Was ist guter Unterricht? Ergebnisse der TIMSS-Videostudie im Fach Mathematik. In: W. Bergsdorf et al. (Eds.), *Herausforderungen der Bildungsgesellschaft* (pp. 89-113). Weimar: Rhino.
- Kuntze, S. (2003). Wie beteiligen Lehrer ihre Schüler an Beweisen im Geometrieunterricht? Erste Ergebnisse einer Auswertung videografiertter Unterrichtsstunden. In W. Henn (Ed.), *Beiträge zum Mathematikunterricht 2003* (pp. 373-376). Hildesheim: Franzbecker.
- Kuntze, S. (2004a). Das binationale und videobasierte Lehrerinnen- und Lehrerfortbildungs-

- projekt „MuBiL“. *GDM-Mitteilungen*, 79, 112-118.
- Kuntze, S. (2004b). Vorstellungen von Mathematiklehrerinnen und -lehrern zur Unterrichtsqualität. In A. Heinze & S. Kuntze (Eds.), *Beiträge zum Mathematikunterricht 2004* (pp. 321-324). Hildesheim: Franzbecker.
- Kuntze, S. (submitted). Zusammenhänge zwischen allgemeinen und situiert erhobenen unterrichtsbezogenen Kognitionen und Überzeugungen von Mathematiklehrerinnen und -lehrern.
- Kuntze, S., Rechner, M. & Reiss, K. (2004) Inhaltliche Elemente und Anforderungsniveau des Unterrichtsgesprächs beim geometrischen Beweisen. Eine Analyse videografiertes Unterrichtsstunden. *matematica didactica*, 27(1), 3-22.
- Kuntze, S., & Reiss, K. (2004). Unterschiede zwischen Klassen hinsichtlich inhaltlicher Elemente und Anforderungsniveaus im Unterrichtsgespräch beim Erarbeiten von Beweisen - Ergebnisse einer Videoanalyse. *Unterrichtswissenschaft*, 32(4), 357-379.
- Kuntze, S., & Reiss, K. (2005). Situation-specific and generalized components of professional knowledge of mathematics teachers – Research on a video-based in-service teacher learning program. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the Int. Group for the Psychology of Mathematics Education (PME)*, Vol. 3 (pp. 225-232). Melbourne: University.
- Leinhardt, G., & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78, 75-95.
- Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde der Forschung und mögliche Konsequenzen für die Praxis. *Die Deutsche Schule*, 96(4), 462-479.
- Lipowsky, F., Thußbas, C., Klieme, E., Reusser, K., & Pauli, C. (2003). Professionelles Lehrerwissen, selbstbezogene Kognitionen und wahrgenommene Schulumwelt – Ergebnisse einer kulturvergleichenden Studie deutscher und Schweizer Mathematiklehrkräfte. *Unterrichtswissenschaft*, 31(3), 206-237.
- Peterson, P., Fennema, E., Carpenter, T., & Loef, M. (1989). Teacher's pedagogical content beliefs in mathematics. *Cognition and Instruction*, 6, 1-40.
- Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik – Ansätze der TIMSS 1999 Videostudie und ihrer schweizerischen Erweiterung. *ZDM*, 35(6), 265-280.
- Reiss, K. (2002). Beweisen, Begründen, Argumentieren. Wege zu einem diskursiven Mathematikunterricht. In W. Peschek (Ed.), *Beiträge zum Mathematikunterricht 2002* (pp. 39-46). Hildesheim: Franzbecker.
- Reiss, K.; Hellmich, F., & Thomas, J. (2002). Individuelle und schulische Bedingungsfaktoren für Argumentationen und Beweise im Mathematikunterricht. In M. Prenzel & J. Doll (Eds.), *45. Beiheft zur Zeitschrift für Pädagogik* (pp. 51-64). Weinheim: Beltz.
- Shulman, L. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 3-36). New York: Macmillan.
- Staub, F., & Stern, E. (2002). The Nature of Teacher's Pedagogical Content Beliefs Matters for Students' Achievement Gains. *Journal of Educational Psychology* 94(2), 344-355.
- Stern, E., & Staub, F. (2000). Mathematik lehren und verstehen: Anforderungen an den Unterricht. In E. Inckermann, J. Kahler, & A. Speck-Hamdan (Eds.), *Sich Lernen leisten*. (pp. 90-100). Neuwied: Luchterhand.
- Tillema, H. (2000). Belief change towards self-directed learning in student teachers: Immersion in practice or reflection on action. *Teaching and Teacher Education*, 16, 575-591.

Author

Sebastian Kuntze
 Lehrstuhl für Didaktik der Mathematik
 Ludwig-Maximilians-Universität München
 Theresienstr. 39, 80333 München
 E-Mail: kuntze@math.lmu.de