



# European Journal of Educational Research


Volume 12, Issue 2, 851 - 863.

ISSN: 2165-8714

<http://www.eu-jer.com/>

## Evaluation of Preservice Teachers' Performance in School through Video Observations during the COVID-19 Pandemic

Maria Carme

Peguera-Carré\* 

University of Lleida, SPAIN

Jordi Coiduras 

University of Lleida, SPAIN

David Aguilar 

University of Lleida, SPAIN

Àngel Blanch 

University of Lleida, SPAIN

Received: September 24, 2022 • Revised: January 6, 2023 • Accepted: March 9, 2023

**Abstract:** Lesson study, observation and analysis are relevant to professional development and initial teacher education. As a strategy, it helps to bridge the gap between theory and practice. The health conditions brought on by the COVID-19 pandemic forced the restriction of the tutors' direct observation of preservice teachers at school. This study analyses preservice teachers' performance through video observations to evaluate their professional activity at school during the COVID-19 pandemic. The Fifteen Items Revised Tsang-Hester Observation Rubric (FIR-THOR) was administered to a sample of 166 preservice teachers in their internship schools and their video recordings each one of 45-minute teaching lessons were analysed. The results show that the FIR-THOR appears as a robust instrument, which allows us to conclude that the instrument works well in the three five-items dimensions that compose it - Instruction, Management, and Assessment - proving to be reliable for assessing teacher intervention in the classroom. Among the three dimensions, the preservice teachers' performance stands out in the Management of the classroom, as well as in the classroom Instruction. This contribution is relevant considering the potential of lesson analysis in learning and professional development during initial teacher training.

**Keywords:** *Observation, preservice teachers, rubrics, student evaluation, teacher education.*

**To cite this article:** Peguera-Carré, M. C., Coiduras, J., Aguilar, D., & Blanch, À. (2023). Evaluation of preservice teachers' performance in school through video observations during the COVID-19 pandemic. *European Journal of Educational Research*, 12(2), 851-863. <https://doi.org/10.12973/eu-jer.12.2.851>

### Introduction

#### *Observation and Evaluation of Preservice Teachers*

One of the main concerns in the training of preservice teachers (PTs) is to progressively guarantee the development of teaching skills. For example, the observation of teaching performance in real situations provides valuable information for monitoring teacher learning and building formative feedback (de Grift et al., 2014; Roegman et al., 2016). It allows understanding PTs' instruction by evaluating high-leverage practices (Ball & Forzani, 2009; McDonald et al., 2013), core practices (Grossman et al., 2009) and approaches to practice (Kazemi et al., 2016). Thus, it informs about the adequacy and consistency of the formative programs in the university and the adaptations that can be agreed upon to better support the teaching practice.

Different studies have been focused on analysing observation practices to collect reliable information regarding different aspects of PTs' teaching practices and their achievements in order to improve instructional quality (Begrich et al., 2021; Lavigne & Good, 2015). Observation protocols and instruments are required to assess specific teaching aspects of instructional intervention in the classroom (Hill et al., 2012; Liu et al., 2017). There are different observation frameworks that allow conceptualising teaching from different perspectives, through generic or content-specific dimensions or with a combination of both (Charalambous & Praetorius, 2018). In the review of Klette and Blikstad-Balas (2018), they analysed coding and observation manuals and detected common generic dimensions: instructional format and clarity, cognitive challenge, evaluation, teachers-student interactions and classroom management. An example of an observation tool is the Tsang-Hester Observation Rubric (THOR), which evaluates PTs' classroom management, instruction, and assessment (Good et al., 2006).

These observations can be translated into a score useful for improvement-oriented feedback in individual teacher evaluation or for research purposes (Bell et al., 2019). For instance, Kraft and Blazar (2017) highlight how the

#### \* Corresponding author:

Maria Carme Peguera-Carré, Department of Education Sciences, University of Lleida, Spain. ✉ [mariacarme.peguera@udl.cat](mailto:mariacarme.peguera@udl.cat)



observations were useful for providing individualised advice to both in-service and PTs as feedback for planning their classroom interventions. Nevgi and Löfström (2015) and Lasagabaster and Sierra (2011) also emphasise that observation should not be detached from classroom reality, encouraging practice reflection which could arise strategies to improve teaching.

#### *Evaluation at Practicum through Video Observations*

Video observations are being increasingly used to analyse teaching and learning situations from the teachers' and students' perspectives at the same time (Fischer & Neumann, 2012).

Video recordings allow knowing in a quite realistic way specific professional situations, such as the development of the lesson in the class, in order to bring the university training closer to the PT's activity. In particular, this resource allows for the formative evaluation of teaching practice and the promotion of the tutor's dialogue and feedback with the PTs (Gaudin & Chaliès, 2015; Richards et al., 2021). Video observations also register non-verbal aspects of interactions, gestures, facial expressions and movements in space that are not observable by other methods (Krug, 2009).

On the one hand, video recordings serve to address one's own lessons or check one's own behaviour to stimulate personal teacher development and reflection on PTs' self-observation (Dalehefte & Kobarg, 2013; van Es et al., 2017). On the other hand, video recordings are advantageous because they can be paused in order to make annotations or rewind to repeat a part for a better understanding. Another advantage is that videos can be evaluated by different observers, PTs, the academic tutor or the school mentor. Moreover, video recordings bring the opportunity of studying the same part with a different analytical approach (Blikstad-Balas & Sørvik, 2015; Jewitt, 2012), and it makes the analyses of teaching situations more precise via smaller intervals (Klette & Blikstad-Balas, 2018).

However, video observations of PTs' performance also have some limitations that should be considered (Blomberg et al., 2013; Chen et al., 2020; Goldman et al., 2007). First, the video data could be limited compared to in situ classroom observations (Sherin, 2003). Moreover, as Krammer et al. (2006) mentioned, one disadvantage of video observations is that it exposes an image of reality that is based on a partial vision of the scene, affected by the focus and the angle of the camera. Furthermore, video is presumed to illustrate an objective view of classroom situations, but the filter of the viewer's background could affect the objectivity of the observational task (Miller & Zhou, 2007).

Overall, the combination of video and video annotations through observations brings the opportunity to review, analyse and reflect on classroom practices to explore the pedagogical development (Gazdag, et al., 2019). Furthermore, the credibility of observations should improve via quality-control strategies, such as reliability testing, member checking and secondary analysis. Video observations supported by coding instruments foster univocity and clarity in the analysis, so it brings renewed interest in observation instruments and their systematic testing of validity (Rich & Hannafin, 2009). Some instruments, such as the Classroom Assessment Scoring System (Pianta et al., 2008) and the Danielson's Observation Scale (Danielson, 1996, 2013; Tournaki et al., 2009), have been already used with video recordings of lessons to assess teaching practice and curriculum development.

The present situation with the COVID-19 pandemic made observations through video become more important because it allowed university tutors to access classrooms without attending in person in order to evaluate PTs' performance (Moyo, 2020). In this context, university tutors faced unforeseen needs and challenges (Hodges et al., 2020), adopting emergency solutions, with the aid of available digital tools (Williamson et al., 2020).

#### *Present Study*

The COVID-19 pandemic and the abrupt decision of having closed groups during 2020-2021 created an assessment necessity of change concerning PTs on practicum. This recent and exceptional situation implies a research gap and few studies have tackled the assessment of PTs during their practicum in the schools. Emergent research has been mainly focused on written feedback or reflective practice and critical discussions via online platforms to follow the practicum (Murtagh, 2022; Prilop et al., 2019).

The objective of the present study was to analyse the PTs' performance through video recordings for assessing their professional activity at school during the pandemic. Thus, this study describes an alternative strategy of assessment to measure the skills that PTs had developed. The proposal of the Fifteen Items Revised Tsang-Hester Observation Rubric (FIR-THOR) (Coiduras et al., 2020) was implemented for evaluating remotely PTs' performance in primary education classrooms. This instrument was chosen because it proposes broad categories that synthesise dimensions of previous proposals: classroom management, instructions and assessment (Bell et al., 2019; Good et al., 2006). Besides, it is a more manageable and efficient tool, which facilitates data collection to a greater extent (Coiduras et al., 2020). This leads us to the research question of the present study:

Do video recordings combined with the FIR-THOR instrument allow to analyse the PTs' performance at school?

## Methodology

This study is focused on the PTs' professional activity at school during the COVID-19 pandemic. Specifically, the FIR-THOR instrument was applied to assess their school performance through videos, and a Confirmatory Factor Analysis (CFA) was used to validate the instrument.

### Participants

The participants in this study were 166 PTs (60,24% females; mean age between 19 and 21 years old) enrolled in the 2020–2021 academic year. They were in their first ( $n = 54$ ), second ( $n = 54$ ) and third year ( $n = 58$ ) of the Primary Education Degree in a dual-system of a southern European University. PTs teachers in the current dual-system attend, from 1st to 4th years of the degree, 40% of the face-to-face activity in schools. In the 1st year, participants of this study develop their practices in urban schools of medium or high socioeconomic status, in the 2nd year they attend rural schools where pupils of different ages are taught together and, in the 3rd year, they attend urban schools with high diversity and located in a disadvantaged socioeconomic context.

### Instrument

The fieldwork for this study was based on the videos recorded by the PTs of one of their lessons during their practicum. PTs needed to fulfil the following conditions and technical aspects of their video recordings, of which they were informed before their performance. They should take into account the location of the camera and its horizontal framing to show all the participants in the class; the image conditions, without backlighting, lack or excess of lighting; the audio, avoiding background sounds and not placing the camera too far away, in order to be able to hear the voices of the transmitters correctly. Finally, it was requested that the recordings should be a sequential shot of approximately 45 minutes, without cuts or edition, and, it was asked to deliver the video in MP4, MOV, WMV or AVI format with a minimum resolution of 720 pixels (1280 x 720).

The video recordings were evaluated with the FIR-THOR instrument by the university tutor. The instrument comprehends three five-items dimensions (Coiduras et al., 2020): Instruction, Management, and Assessment (Appendix). The instrument was translated into Catalan language bearing in mind the equivalence of item meaning (Blanch & Aluja, 2016) and showed generally robust reliability and validity psychometric properties (Coiduras et al., 2020). The three five-items dimensions were assessed on a five-point Likert scale. Instruction addresses the implemented strategies and activities, the quality and appropriate use of teaching resources, the questions for cognitive activation, the connection of pupils with the content for a deeper understanding, and the PTs' motivation when conducting the lesson. The Management dimension relates to the classroom environment, participation and student engagement that favours a productive learning activity. The Assessment dimension taps into the planning of the pupils' evaluation and the communication in the classroom of the lesson objectives and the evaluation criteria.

### Data Collection and Scoring

The video recordings were analysed by four university tutors, two of them authors of the present paper, during 2020-2021. These tutors previously expressed their agreement with the type of analysis to be performed and also to participate in the study. At that time, schools restricted visitors because of the COVID-19 pandemic. Thus, tutors collected school-based material for the assessment of the PTs' competence through video recordings. PTs and tutors signed the data protection and legal authorisation document for the recording of children. After obtaining informed consent and image rights as university students and with the school authorisation, PTs were recorded on video while performing 45-minute teaching lessons to primary education students. Figure 1 presents an excerpt of the initial part of one of the PTs' video-recorded lessons at a school.



Figure 1. Excerpt of a PT's Lesson Video Recording (Peguera, 2022)

In order to be included as a participant in the present study, PTs needed to fulfil the following criteria: (a) submission of a video of a lesson conducted in the primary education classroom and the corresponding syllabus design; (b) appropriate technical conditions of the video recording in terms of sound and image; (c) acting in the classroom intervention as the only teacher and without their mentor assistance. In the end, the video recordings of 166 different PTs were included in the current study.

To analyse the PTs' performance, all lesson video recordings were subdivided into three equal intervals of approximately 15 minutes each (Figure 2). Each interval was analysed with the FIR-THOR instrument (Appendix). Therefore, all items of the three five-items dimensions were assessed in all the three intervals to discriminate possible fluctuations during the PTs teaching practice.

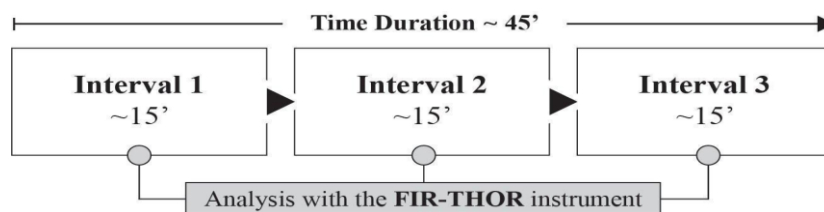


Figure 2. Video Recording Evaluation Process with the FIR-THOR Instrument at Three Observation Intervals

The four tutors, all of them university lecturers involved in PTs' tutoring and pedagogical research, were trained for the scoring of the test task. To reach an interrater consensus for the assessment of the video recordings using the Appendix instrument, an iterative process with successive rounds was performed. Using videos of the sample with a variety of course levels and subjects taught, the four researchers discussed the evaluation of the items and, in the last round, an agreement superior to .80 was reached. Where there was some discrepancy, they discussed the meaning of that item to reach a total agreement.

#### Data Analyses

The mean scores across the three observation intervals were evaluated with an ANOVA repeated measures approach ( $F$ -test). Post-hoc Tukey multiple comparisons of the mean scores at each interval were additionally estimated to contrast the specific interval points with statistically significant differences (Tukey, 1949). The null hypothesis here was that the means across the three observation intervals were equal.

A CFA was conducted for the data obtained with the FIR-THOR instrument in each of the three observation intervals shown in Figure 2 (Bollen, 1989). This model was specified with three correlated latent variables corresponding to the Instruction, Management, and Assessment dimensions (Coiduras et al., 2020). Each of the three latent factors was measured with their corresponding five items, Preliminary analyses suggested an abnormally high modification index concerning the correlation between items 1 (A1) and 2 (A2) in the Assessment dimension. Therefore, all analyses were conducted with both items being specified as correlated (Blanch & Aluja, 2009; Byrne, 2001). Normality assumptions were met by all items in the evaluated instrument, with skewness and kurtosis values lower than 1. The only exception was item 1 in the management dimension, which showed two outliers that would be unlikely to alter the main outcomes of the CFA conducted in the current study.

Model evaluation was conducted with the  $\chi^2$  value, the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR). Low and non-significant values in the  $\chi^2$ , values above .90 in the CFI and the TLI, and values below .09 in the RMSEA and the SRMR are generally considered supportive of an acceptable fit to the observed data (Hu & Bentler, 1999). The three CFA models were estimated with the maximum likelihood method from the lavaan package R software (R Development Core Team, 2014; Rosseel, 2012).

## Results

#### Descriptive Analyses

Table 1 and Appendix show the descriptive statistics for the Instruction, Management, and Assessment items and dimensions across the three observation intervals, obtained with the new FIR-THOR instrument. Cronbach's alpha internal consistencies were fair for Instruction and Assessment and high for Management.

Table 1. Mean, Standard Deviations, Cronbach's Alpha Internal Reliability ( $\alpha$ ), and ANOVA of Repeated Measures

	Interval 1			Interval 2			Interval 3			<i>F</i>
	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$	<i>M</i>	<i>SD</i>	$\alpha$	
Instruction	18.78	4.23	.83	18.10	4.44	.87	17.48	4.57	.86	3.61*
Management	21.84	3.66	.88	21.04	3.99	.91	20.23	4.70	.93	6.32**
Assessment	16.81	4.29	.82	16.45	4.33	.82	16.52	4.28	.80	.32

\* $p < .05$ , \*\* $p < .01$ .

The mean scores in the three dimensions decreased from Interval 1 to Interval 3. An analysis of variance for repeated measures (ANOVA) indicated that these decrements were statistically significant for the Instruction ( $F(2,495) = 3.61, p = .0279$ ) and Management dimensions ( $F(2,495) = 6.32, p = .0019$ ), but non-significant for the Assessment dimension ( $F(2,495) = .32, p = .7260$ ). The Tukey multiple comparisons of means indicated a significant difference of 1.30 for Instruction (95% CI = [.16, 2.44],  $p = .0204$ ), and a significant difference of 1.61 for Management (95% CI = [.55, 2.68],  $p = .0012$ ).

*Confirmatory Factor Analysis*

Figure 3 shows the CFA for the FIR-THOR instrument at three observation intervals. The correlations amongst the latent variables were also highly significant ( $p < .001$ ). Nonetheless, these correlations were larger between the Instruction and Assessment dimensions (.83, .84, and .82). The correlations between the residual terms of items 1 (A1) and 2 (A2) in the Assessment dimension were also highly significant (.36, .40, .33,  $p < .001$ ). The standardised factor loadings were also highly significant.

Table 2 shows the fit indices of the models obtained at Interval 1, Interval 2, and Interval 3. The evaluation of these models highlights a fair model fit throughout the three observation intervals. Despite the statistical significance of the three chi-square values, the CFI, TLI, RMSEA, and SRMR values ranged within acceptable standards (Hu & Bentler, 1999), with a somewhat evident deterioration in model fit from Interval 1 to Interval 3. The FIR-THOR instrument, however, appears as a robust instrument to evaluate teacher performance observed through video recordings.

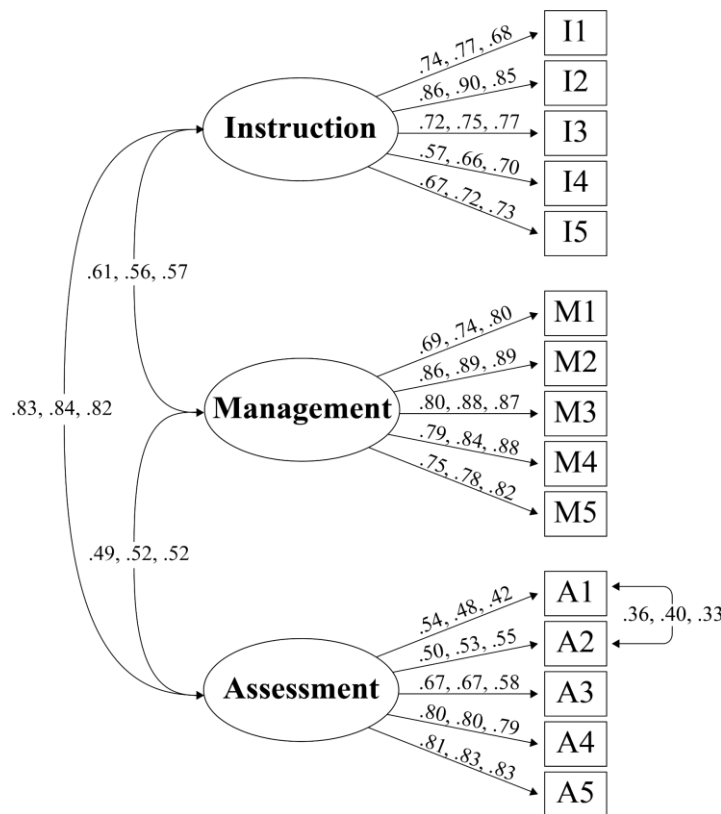


Figure 3. CFA of the FIR-THOR Instrument at Three Observation Intervals. Standardised Factor Loadings and Latent Factor Correlations were Significant at the  $p < .001$  Level (The Three Values Indicate the Correlations and Factor Loadings at Interval 1, Interval 2, and Interval 3, respectively)

Table 2. Fit Indices of the CFA Model at the Three Evaluation Moments (Interval 1, Interval 2, and Interval 3). There were 86 Degrees of Freedom for Each of the Three CFAs, and the Three Chi-Square Values were Significant at the  $p < .001$  Level

Fit index	Interval 1	Interval 2	Interval 3
$\chi^2$	160.18	171.80	205.15
CFI	.929	.932	.905
TLI	.914	.918	.896
RMSEA	.080	.086	.103
SRMR	.077	.068	.082

## Discussion

Classroom lesson observation has emerged over the last decades as a pivotal tool for improving the professional skills and knowledge base of PTs and in-service teachers at all stages of their careers (O'Leary, 2012, 2013, 2020). The pandemic situation made it impossible to obtain substantial and realistic information on the achievements of the PTs for their evaluation and consequent feedback. The combination of video recording and evaluation with FIR-THOR (Table 2 and Figure 3) has provided an emergency solution with a reliable response to the encountered constraints (Hodges et al., 2020).

In this study, the PTs' performance was analysed through video recordings evaluated with the FIR-THOR instrument. The obtained results suggest that video observation combined with this instrument can be implemented as an alternative evaluation strategy in the initial teacher training.

Observations through video recordings provided a digital approach to the PTs' assessment process. Video recordings had helped the tutors to return, if needed, multiple times to moments of interest of the PTs' performance when difficulties or doubts arose during the evaluation. This allowed the observation of the same scene several times to analyse in more detail how the events unfolded in the teaching performance (Spiro et al., 2007). Comparatively, *in vivo* observations tend to be more complicated since the tutor's attention is divided, looking alternately and not simultaneously at the PT and the observation grid. Videos facilitated the viewing by pausing the action in order to make a more precise annotation. However, digital images provide a fixed point of view which leads to a loss of contextual information about the classroom lesson in comparison to *in vivo* observations. Recordings were mainly focused on the PTs' performance, but actions that took place outside of the video scene, i.e., interactions between pupils were missed since there was a limitation to capture everything that happened in the classroom (Blomberg et al., 2013; Krammer et al., 2006; Sherin, 2003). This limitation could be reduced in some situations by using more than one camera to record the lesson or by giving background information (Begrich et al., 2021; Miller & Zhou, 2007).

Results from the implementation of the FIR-THOR showed it as a robust strategy to evaluate PTs performance (Figure 3 and Table 2), even when used under the pandemic exceptional circumstances. With the three five-items dimensions proposed by Coiduras et al. (2020) the observation is focused on a set of main aspects of the teacher performance which facilitates its use, even by students. This study confirmed that the instrument was able to represent the three teaching dimensions and it appeared as valid and reliable to assess PTs' teaching performance through video recordings.

Concerning the PTs' achievements and performance, there were medium to high scores on the three dimensions during the three assessed intervals (Table 1).

Management consists of creating a scenario favourable to learning, of positive coexistence, and of relationships based on respect and of the student's commitment to the activity (Bear, 2015; Danielson, 2013; Evertson & Weinstein, 2006; Yang et al., 2018). Management yielded the highest score during the three observed intervals ( $M = 21.84, 21.04, 20.23$ , respectively). It was observed that PTs created safe and stimulating learning environments, with a positive climate of mutual respect and supportive communication. For example, the following transcribed excerpt of one of the videos evidences these aspects:

PT 1: Today we will test all that we have learned and what we have just seen that we remember quite well. [Pupil 1 raises the hand] Yes?

Pupil 1: I have a Conan comic book.

PT 1: That's good. And do you like reading comics?

Pupil 1: Yes.

PT 1: More than novels? Why?

Pupil 1: Because they have more action, more drawings...

PT 1: All right. So look, today we will make our own comic book with a picture that you will take from this box. For example, this picture is from Magnolia park.

*(Excerpt from a Video Sample #1)*

In this situation the PT attended to pupils' participation, encouraged dialogue and took advantage of their sporadic interventions to draw out further concepts about the taught topic. This also helped pupils to feel confident and comfortable. Good et al. (2006) highlight that, under these conditions, PTs tend to have a good relationship with pupils and, thus, favourable learning environments are promoted.

The Instruction dimension assesses the way PTs conducted the teaching and supported the learning process (Baumert et al., 2010; Osborne et al., 2015). This dimension scored a bit lower than Management in all the three assessed intervals ( $M = 18.78, 18.10, 17.48$ , respectively). In the Instruction dimension, PTs showed the highest scores when showing their conviction and energy during the lesson. Probably, the fact of being observed and evaluated could act as a further catalyst

for promoting PT's motivation (Smith & Coombs, 2003). Other assessed items related to the content knowledge, the formulation of questions, the use of examples, models or visual representations showed room for improvement. Most of the observed PTs showed a lack of in-depth understanding when introducing new or difficult concepts in agreement with their short teaching experience. As Ball et al. (2008) argue "teachers who do not know a subject well are not likely to have the knowledge they need to help students learn this content" (p. 404). They ask relevant questions with an instructional intent, but with a formulation that does not mostly promote thoughtful responses, which is probably in consonance with the observed lack of in-depth content knowledge (Venkat et al., 2014). For example, the following transcribed excerpt of one of the videos evidences the formulation of questions:

PT 2: What is the process of getting a chick out of an egg called?

Pupils: Incubation.

PT 2: Incubation, very good, and what are we going to use for incubation?

Pupils: The incubator.

PT 2: An incubator. All right, so now, I'm going to hand you the incubator, you will observe it and tell me what features you see of the incubator.

*(Excerpt from a Video Sample #2)*

In this excerpt the PT tries to promote pupils' ideas related to a science topic, although the formulated factual questions only encourage, at first, short answers. Enhancing the core practice of questioning during the initial teacher training could help PTs to provide pupils with more opportunities to further understand the taught content (Bell et al., 1985; Hackenberg, 2005; Harlen, 2001).

The classroom Assessment items relate to the review of previous learning, verification of understanding and the kind of feedback provided to pupils to improve learning (Black & Wiliam, 1998, 2010). This dimension showed the lowest scores during the three observed intervals (M = 16.81, 16.45, 16.52, respectively). These results could be explained by a low systematisation of the assessment tasks in comparison with the other observed dimensions. The assessed videos highlight that PTs plan a set of criteria and standards for evaluating their lessons, but with a lack of clear guidelines and rubrics. It is also observed that they use to provide general feedback without answering the individual needs of the pupils in the classroom. As Acar-Erdol and Yıldızlı (2018) and Herzog-Punzenberger et al. (2020) remark, this kind of assessment process only provides a partial estimation of the pupils' learning achievements, which could hinder the adaptation of the teaching tasks to the diversity found in class.

It is interesting to highlight the evolution of the scores in each dimension of the lesson. Scores dropped significantly from Interval 1 to 3 in Instruction and Management, but this decrease was not significant in Assessment (Table 1). Regarding Management this could be explained by a decrease in the PTs' communication ability with pupils in moments of autonomous work dynamics and implementation of strategies. Also, in some cases, a decrease in attention has been observed, which means tiredness and fatigue appeared during prolonged teaching sessions. This could explain this progressive deterioration of the scores in the Management dimension. Regarding Instruction, we have observed that in Interval 3 the PTs end the lesson with a synthesis-evaluation of the learning achieved and of the processes and dynamics. In some cases, the session ends with explanations that do not directly connect with the importance of the content and its relationship with other subjects previously worked on, which has resulted in lower scores in the last interval of the Instruction dimension.

### Conclusion

The purpose of this study was to analyse the PTs' performance through video recordings of their professional activity at school during the pandemic. PTs evaluation combining video recordings with the FIR-THOR has been proven to be a valid and reliable strategy for teacher training. The results showed that PTs had high scores on the Management and Instruction dimensions. Although PTs scored well also on the Assessment items, this study highlights that the use of classroom assessment strategies will continue to be a challenging task in teacher preparation programs (Battistone et al., 2019). Thus, video observations allowed us to study the dimensions and elements involved in teaching to obtain evaluative feedback. Besides, video observations with the FIR-THOR offered the study new possibilities for analysing experiences in school classrooms in relation to planning, teaching and group management.

### Recommendations

The outcomes from this study highlight that the video approach made PTs' observations more flexible and not limited by distance, space, and time. Video observation helped tutors to assess PTs' performance under the exceptional circumstances derived from the pandemic.

The combination of the FIR-THOR and video is a valid strategy for analysing and studying lessons performed at school. It contributes to the quality of the practicum by providing PTs with learning environments in which they have the opportunity to reflect on theory and practice simultaneously. Therefore, there is a potential to be explored in initial and



permanent training. This analysis strategy can contribute to the theoretical component of knowledge based on research that is needed in teacher training. Moreover, the derived analyses open the opportunity to review the instructional sequence, in an exercise that helps to reflect on the planning and intervention already performed in the classroom. As Kramer et al. (2020) and Zaragoza et al. (2021) indicate, teaching programmes need to take into account this connection of theoretical knowledge with practice.

The use of these tools promotes reflection and feedback through dialogue between PTs and the different agents involved in teacher education. Feedback from mentors and peers might be beneficial for the teaching procedures' awareness and self-regulation, which, as Ronen (2022) highlighted, are relevant to professional teacher development.

Finally, PTs should be encouraged to use different tools to support their observations and reflections on their practice. Observations based on criteria of teaching performance should support them to identify strengths and explicit opportunities for improvement to be addressed in a constructive conversation, focusing on the specific performance in real life.

### Limitations

Although the instrument proved its consistency, the study has a reduced sample size limited by the COVID-19 pandemic. Another limitation is that the participants of this study come from a specific region and not an international study. Future studies may be done in collaboration with other countries or research settings to obtain more generalised data about lesson instruction' written reflections.

### Acknowledgements

The authors are grateful to the school principals, teachers, parents and pupils, and also the preservice teachers, for their permission and assistance to collect the data in the classrooms during the COVID-19 pandemic.

### Funding

The authors would like to acknowledge the support received for this research (2017 ARMIF 00028 and 2020 ARMIF 00019) from the Agència d'Ajuts per a la Recerca i Universitats (AGAUR) and from the Catalan Departament d'Economia i Societat del Coneixement, as well as the help of the University of Lleida in the Program of Research Promotion 2019. This research was performed within the Competences, Technology and Society in Education Research Group (COMPETECS, 2021 SGR 01360).

### Authorship Contribution Statement

Peguera-Carré: Conceptualisation, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation, Project Administration. Coiduras: Conceptualisation, Investigation, Resources, Writing - Review & Editing, Supervision, Project Administration, Funding acquisition. Aguilar: Conceptualisation, Resources, Writing - Review & Editing, Supervision, Funding acquisition. Blanch: Methodology, Software, Validation, Formal analysis, Writing - Review & Editing.

### References

- Acar-Erdol, T., & Yıldızlı, H. (2018). Classroom assessment practices of teachers in Turkey. *International Journal of Instruction*, 11(3), 587-602. <https://doi.org/10.12973/iji.2018.11340a>
- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497-511. <https://doi.org/10.1177/0022487109348479>
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. <https://doi.org/10.1177/0022487108324554>
- Battistone, W., Buckmiller, T., & Peters, R. (2019). Assessing Assessment Literacy: Are new teachers prepared to assume jobs in school districts engaging in grading and assessment reform efforts? *Studies in Educational Evaluation*, 62, 10-17. <https://doi.org/10.1016/j.stueduc.2019.04.009>
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133-180. <https://doi.org/10.3102/0002831209345157>
- Bear, G. G. (2015). Preventive and classroom-based strategies. In E. T. Emmer, & E. J. Sabornie (Eds.), *Handbook of classroom management* (pp. 15-39). Routledge.
- Begrich, L., Kuger, S., Klieme, E., & Kunter, M. (2021). At a first glance—How reliable and valid is the thin slices technique to assess instructional quality? *Learning and Instruction*, 74, Article 101466. <https://doi.org/10.1016/j.learninstruc.2021.101466>
- Bell, B., Osborne, R., & Tasker, R. (1985). Finding out what children think. In R. Osborne, & P. Freyberg (Eds.), *Learning in*



- science: *The implications of children's ideas* (pp. 151–165). Heinemann.
- Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School Effectiveness and School Improvement*, 30(1), 3-29. <https://doi.org/10.1080/09243453.2018.1539014>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. <https://doi.org/10.1177/003172171009200119>
- Blanch, A., & Aluja, A. (2009). Validation study of the Spanish Version of the Work-Family Conflict Questionnaire (CCTF). *Spanish Journal of Psychology*, 12(2), 746-755. <https://doi.org/10.1017/S1138741600002110>
- Blanch, A., & Aluja, A. (2016). Text mining a self-report back-translation. *Psychological Assessment*, 28(6), 750–764. <https://doi.org/10.1037/pas0000213>
- Blikstad-Balas, M., & Sørvik, G. O. (2015). Researching literacy in context: Using video analysis to explore school literacies. *Literacy*, 49(3), 140-148. <https://doi.org/10.1111/lit.12037>
- Blomberg, G., Renkl, A., Gamoran Sherin, M., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online*, 5(1), 90–114. <https://bit.ly/4280gkL>
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons. <https://doi.org/10.1002/9781118619179>
- Byrne, B. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. Lawrence Erlbaum.
- Charalambous, C. Y., & Praetorius, A. K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM - Mathematics Education*, 50, 355–366. <http://doi.org/10.1007/s11858-018-0914-8>
- Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *Journal of the Learning Sciences*, 29(4-5), 642-680. <https://doi.org/10.1080/10508406.2020.1783269>
- Coiduras, J. L., Blanch, À., & Barbero, I. (2020). Initial teacher education in a dual-system: Addressing the observation of teaching performance. *Studies in Educational Evaluation* 64, Article 100834. <https://doi.org/10.1016/j.stueduc.2019.100834>
- Dalehefte, I. M., & Kobarg, M. (2013). *Aus Unterrichtsbeobachtungen lernen: Handreichung des Programms SINUS an Grundschulen* [Learning from lesson observations: Handout of the SINUS programme at primary schools]. IPN – Leibniz Institute for Science and mathematics Education at Kiel University.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Association for Supervision and Curriculum Development.
- Danielson, C. (2013). *The Framework for Teaching Evaluation Instrument*. Danielson Group.
- de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: Calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation*, 43, 150-159. <https://doi.org/10.1016/j.stueduc.2014.09.003>
- Evertson, C. M., & Weinstein, C. S. (Eds.). (2006). *Handbook of classroom management: Research, practice, and contemporary issues*. Routledge. <https://doi.org/10.4324/9780203874783>
- Fischer, H. E., & Neumann, K. (2012). Video analysis as a tool for understanding science instruction. In D. Jord, & J. Dillon (Eds.), *Science education research and practice in Europe: Cultural perspectives in science education* (Vol. 5, pp. 115-139). Sense Publishers. [https://doi.org/10.1007/978-94-6091-900-8\\_6](https://doi.org/10.1007/978-94-6091-900-8_6)
- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gazdag, E., Nagy, K., & Szivák, J. (2019). "I Spy with My Little Eyes..." The use of video stimulated recall methodology in teacher training–The exploration of aims, goals and methodological characteristics of VSR methodology through systematic literature review. *International Journal of Educational Research*, 95, 60-75. <https://doi.org/10.1016/j.ijer.2019.02.015>
- Goldman, R., Pea, R., Barron, B., & Derry, S. J. (Eds.). (2007). *Video research in the learning sciences*. Lawrence Erlbaum.
- Good, T., McCaslin, M., Tsang, H., Zhang, J., Wiley, C., Bozack, A., & Hester, W. (2006). How well do 1st-year teachers teach

- does type of preparation make a difference? *Journal of Teacher Education*, 57(4), 410-430. <https://doi.org/10.1177/0022487106291566>
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055-2100. <https://doi.org/10.1177/016146810911100905>
- Hackenberg, A. (2005). A model of mathematical learning and caring relations. *For the Learning of Mathematics*, 25(1), 44-47. <https://www.jstor.org/stable/40248486>
- Harlen, W. (Ed.). (2001). *Primary science ...taking the plunge: How to teach primary science more effectively for ages 5 to 12* (2nd ed.). Heinemann.
- Herzog-Punzenberger, B., Altrichter, H., Brown, M., Burns, D., Nortvedt, G. A., Skedsmo, G., Wiese, E., Nayir, F., Fellner, M., McNamara, G., & O'Hara, J. (2020). Teachers responding to cultural diversity: case studies on assessment practices, challenges and experiences in secondary schools in Austria, Ireland, Norway and Turkey. *Educational Assessment, Evaluation and Accountability*, 32, 395-424. <https://doi.org/10.1007/s11092-020-09330-y>
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56- 64. <https://doi.org/10.3102/0013189X12437203>
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). *The difference between emergency remote teaching and online learning*. EDUCAUSE Review. <http://hdl.handle.net/10919/104648>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Jewitt, C. (2012). *An introduction to using video for research* (NCRM Working Paper). National Centre for Research Methods. <http://eprints.ncrm.ac.uk/2259/>
- Kazemi, E., Ghouseini, H., Cunard, A., & Turrou, A. C. (2016). Getting inside rehearsals: Insights from teacher educators to support work on complex practice. *Journal of Teacher Education*, 67(1), 18-31. <https://doi.org/10.1177/0022487115615191>
- Klette, K., & Blikstad-Balas, M. (2018). Observation manuals as lenses to classroom teaching: Pitfalls and possibilities. *European Educational Research Journal*, 17(1), 129-146. <https://doi.org/10.1177/1474904117703228>
- Kraft, M. A., & Blazar, D. (2017). Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Educational Policy*, 31(7), 1033-1068. <https://doi.org/10.1177/0895904816631099>
- Kramer, C., König, J., Strauss, S., & Kaspar, K. (2020). Classroom videos or transcripts? A quasi-experimental study to assess the effects of media-based learning on pre-service teachers' situation-specific skills of classroom management. *International Journal of Educational Research*, 103, Article 101624. <https://doi.org/10.1016/j.ijer.2020.101624>
- Krammer, K., Ratzka, N., Klieme, E., Lipowsky, F., Pauli, C., & Reusser, K. (2006). Learning with classroom videos: Conception and first results of an online teacher training program. *ZDM - Mathematics Education*, 38, 422-432. <https://doi.org/10.1007/BF02652803>
- Krug, M. (2009). "Videobasierte Methoden der Bildungsforschung: Sozial-, erziehungs- und kulturwissenschaftliche Nutzungsweisen"(19./20.06.2009, Stiftung Universität Hildesheim): Tagungsbericht. ["Video-based methods in educational research: Ways of using them in social, educational and cultural sciences"(19/20.06.2009, Foundation University of Hildesheim): Conference report.] *Zeitschrift für Qualitative Forschung*, 10(1), 161-167. <https://124.im/7QEwr>
- Lasagabaster, D., & Sierra, J. M. (2011). Classroom observation: Desirable conditions established by teachers. *European Journal of Teacher Education*, 34(4), 449-463. <https://doi.org/10.1080/02619768.2011.587113>
- Lavigne, A. L., & Good, T. L. (2015). *Improving teaching through observation and feedback: Beyond state and federal mandates*. Routledge. <https://doi.org/10.4324/9781315777030>
- Liu, S., Bell, C. A., & Jones, N. D. (2017, March 1-4). *The validity of classroom observation systems in research and applied contexts* [Conference presentation]. Annual Spring meeting of the Society for Research on Educational Effectiveness (SREE), Washington, DC.
- McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education*, 64(5), 378-386. <https://doi.org/10.1177/0022487113493807>

- Miller, K., & Zhou, X. (2007). Learning from classroom video: What makes it compelling and what makes it hard. In R. Goldmann, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 321–334). Lawrence Erlbaum.
- Moyo, N. (2020). COVID-19 and the future of practicum in teacher education in Zimbabwe: Rethinking the ‘new normal’ in quality assurance for teacher certification. *Journal of Education for Teaching*, 46(4), 536-545. <https://doi.org/10.1080/02607476.2020.1802702>
- Murtagh, L. (2022). Remote tutor visits to practicum settings and the changing dynamics between university tutors, school-based mentors and pre-service teachers. *Journal of Further and Higher Education*, 46(3), 354-367. <https://doi.org/10.1080/0309877X.2021.1945557>
- Nevgi, A., & Löfström, E. (2015). The development of academics’ teacher identity: Enhancing reflection and task perception through a university teacher development programme. *Studies in Educational Evaluation*, 46, 53-60. <https://doi.org/10.1016/j.stueduc.2015.01.003>
- O’Leary, M. (2012). Exploring the role of lesson observation in the English education system: A review of methods, models and meanings. *Professional Development in Education*, 38(5), 791–810. <https://doi.org/10.1080/19415257.2012.693119>
- O’Leary, M. (2013). Expansive and restrictive approaches to professionalism in FE colleges: The observation of teaching and learning as a case in point. *Research in Post-Compulsory Education*, 18(4), 348–364. <https://doi.org/10.1080/13596748.2013.847164>
- O’Leary, M. (2020). *Classroom observation: A guide to the effective observation of teaching and learning* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315630243>
- Osborne, J., Berson, E., Borko, H., Busch, K. C., Zaccarelli, F. G., Million, S., & Tseng, A. (2015, August 25-29). *Assessing the quality of classroom discourse in science classrooms*. [Conference presentation]. 16th biennial conference of the European Association for Research in Learning and Instruction, Limassol, Cyprus.
- Peguera, M. C. (2022, January 31). *Video 1. Extract of a video recorded PT’s lesson* [Video]. YouTube. <https://youtu.be/UOLjcoSSLk>
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system™: Manual K-3*. Paul H Brookes Publishing.
- Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2019). How digital reflection and feedback environments contribute to pre-service teachers’ beliefs during a teaching practicum. *Studies in Educational Evaluation*, 62, 158-170. <https://doi.org/10.1016/j.stueduc.2019.06.005>
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Rich, P. J., & Hannafin, M. (2009). Video annotation tools technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education* 60(1), 52–67. <https://doi.org/10.1177/0022487108328486>
- Richards, J., Altshuler, M., Sherin, B. L., Sherin, M. G., & Leatherwood, C. J. (2021). Complexities and opportunities in teachers’ generation of videos from their own classrooms. *Learning, Culture and Social Interaction*, 28, Article 100490. <https://doi.org/10.1016/j.lcsi.2021.100490>
- Roegman, R., Goodwin, A. L., Reed, R., & Scott-McLaughlin, R. M. (2016). Unpacking the data: An analysis of the use of Danielson’s (2007) Framework for Professional Practice in a teaching residency program. *Educational Assessment, Evaluation and Accountability*, 28, 111-137. <https://doi.org/10.1007/s11092-015-9228-3>
- Ronen, I. K. (2022). The experience of teaching: Beyond teaching skills the case of the academy-class practice model. *Studies in Educational Evaluation*, 72, Article 101115. <https://doi.org/10.1016/j.stueduc.2021.101115>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Sherin, M. G. (2003). New perspectives on the role of video in teacher education. In J. Brophy (Ed.), *Using video in teacher education* (Vol. 10, pp. 1–27). Emerald Group Publishing Limited. [https://doi.org/10.1016/S1479-3687\(03\)10001-6](https://doi.org/10.1016/S1479-3687(03)10001-6)
- Smith, I. D., & Coombs, S. J. (2003). The Hawthorne effect: Is it a help or a hindrance in social science research? *Change*, 6(1), 97–111. <https://bit.ly/3ZRZQxd>
- Spiro, R. J., Collins, B. P., & Ramchandran, A. (2007). Reflections on a post-Gutenberg epistemology of video use in ill-structured domains: Fostering complex learning and cognitive flexibility. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 93–100). Lawrence Erlbaum.

- Tournaki, N., Lyublinskaya, I., & Carolan, B. V. (2009). Pathways to teacher certification: Does it really matter when it comes to efficacy and effectiveness? *Action in Teacher Education*, 30(4), 96-109. <http://doi.org/10.1080/01626620.2009.10734455>
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99-114. <https://doi.org/10.2307/3001913>
- van Es, E. A., Cashen, M., Barnhart, T., & Auger, A. (2017). Learning to notice mathematics instruction: Using video to develop preservice teachers' vision of ambitious pedagogy. *Cognition and Instruction*, 35(3), 165-187. <http://doi.org/10.1080/07370008.2017.1317125>
- Venkat, H., Rollnick, M., Loughran, J., & Askew, M. (Eds.). (2014). *Exploring mathematics and science teachers' knowledge: Windows into teacher thinking*. Routledge. <https://doi.org/10.4324/9781315883090>
- Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, 45(2), 107-114. <https://doi.org/10.1080/17439884.2020.1761641>
- Yang, C., Bear, G. G., May, H., & Curby, T. (2018). Multilevel associations between school-wide social-emotional learning approach and student engagement across elementary, middle, and high schools. *School Psychology Review* 47(1), 45-61. <https://doi.org/10.17105/SPR-2017-0003.V47-1>
- Zaragoza, A., Seidel, T., & Hiebert, J. (2021). Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation. *European Journal of Teacher Education*. Advance online publication. <https://doi.org/10.1080/02619768.2021.1996558>

### Appendix

Dimensions and items evaluated by FIR-THOR (Coiduras et al., 2020) and PTs' mean and standard deviation results. This instrument had been adapted from Good et al. (2006).

Dimension	Item	Mean (SD)		
		Interval 1	Interval 2	Interval 3
Instruction	I1. Makes effective use of learning materials to achieve learning goals	3.78 (1.16)	3.60 (1.19)	3.46 (1.16)
	I2. Demonstrating effective "bag of tricks" in presenting new or difficult concepts	3.57 (1.07)	3.37 (1.09)	3.23 (1.13)
	I3. Demonstrates content knowledge in instruction	3.73 (1.12)	3.66 (1.11)	3.51 (1.14)
	I4. Displays energy and conviction for the content being taught	4.13 (0.92)	4.02 (0.95)	3.90 (1.05)
	I5. Quality of questions	3.57 (1.15)	3.46 (1.12)	3.37 (1.23)
Management	M1. Teacher interaction with students	4.57 (0.77)	4.50 (0.77)	4.37 (0.90)
	M2. Student interactions with other students	4.40 (0.83)	4.23 (0.86)	4.09 (1.02)
	M3. Management of instructional groups and Individuals	4.36 (0.89)	4.16 (0.99)	3.96 (1.14)
	M4. Appropriate behaviour is understood and followed by students	4.27 (0.96)	4.07 (1.05)	3.90 (1.14)
	M5. Monitors student behaviour and provides feedback	4.24 (0.97)	4.08 (0.96)	3.91 (1.10)
Assessment	A1. Formal assessment criteria and standards	2.83 (1.19)	2.75 (1.17)	2.89 (1.19)
	A2. Use of formative assessment	3.12 (1.22)	3.09 (1.21)	3.16 (1.27)
	A3. Learning goals for students	3.69 (1.15)	3.56 (1.17)	3.62 (1.15)
	A4. Providing in-class feedback and informal assessment to students	3.53 (1.02)	3.43 (1.05)	3.41 (1.03)
	A5. Fairness and consistency of formal and/or informal assessment	3.64 (1.03)	3.61 (1.05)	3.44 (1.09)