Organisation for Economic Co-operation and Development

DOCUMENT CODE

For Official Use **English - Or. English**

The TALIS Video Study Observation System

This document was prepared by the TALIS Video Study (TVS) International Consortium for the Organisation for Economic Co-Operation and Development (OECD). It describes the observation system of the TALIS Video Study. It provides an overview of the main design features of the two types of observation codes, the six domains of teaching which were measured and the quality control rating processes of the observation system. Additional detail will be available in the TVS technical report that will be released in late 2020.

The lead authors of the codes and associated training materials are Courtney A. Bell (ETS/University of Wisconsin), Yi Qi (ETS), Margaret W. Witherspoon (ETS), Heather Howell (ETS), and Mariana B. Torres (University of California-Los Angeles). Co-developers are documented in the TVS technical report.

Anna Pons, anna.pons@oecd.org

TALIS Video Study Observation System

Background

1. Around the world, researchers, policy makers, parents and children agree that teachers matter. But, we are only just beginning to understand what, how and why teachers do what they do with students. By directly observing teaching in the classroom, the TALIS Video Study is trialling new research methods to shed light on these questions, which are critical to improving education (see Box 1).

2. This paper describes the observation system of the TALIS Video Study. It provides an overview of the main design features of the two types of observation codes, the six domains of teaching which are measured and the quality control rating processes of the observation system.

3. The annexes contain the main materials used in the rating of the videos collected in the TALIS Video Study:

- Annex A contains the training manual developed for raters of the TALIS Video Study. It explains the procedures to be followed in the rating of indicators and components.
- Annex B presents an abridged version of the code and guidance notes for rating observations.
- Annex C provides the rater agreement metrics in quality control processes and main study rating.

4. Further information of the observation system will become available in the Technical Report of the TALIS Video Study.

Box 1. What is the TALIS Video Study

The TALIS Video Study is an innovative study that trials new methodologies to deepen our understanding of teaching and learning at an international scale. To have a more rounded picture of the classroom, the TALIS Video Study collects observation and artefact evidence in addition to survey and achievement data. The TALIS Video Study is purported to:

- understand which aspects of teaching are related to student learning and student non-cognitive outcomes
- observe and document how the teachers from participating countries and economies in the study teach
- explore how various teaching practices are inter-related, and how contextual aspects of teaching are related to the student and teacher characteristics.

The main design features are:

- Common evaluation method: Unlike many studies of teaching and learning, the TALIS Video Study draws on multiple measures of teaching to provide a more rounded picture of practice. The study develops a new set of instruments for evaluating video-recorded teaching practices and classroom artefacts, and fields them consistently in all participating countries and economies.
- Common topic for evaluation: The TALIS Video Study focuses on the teaching and learning of a single common secondary mathematics topic (quadratic equations) to enhance the comparability across countries and the potential to capture the relationship between teaching and student outcomes. Using mathematics helps to reduce potential differences between countries in terms of curriculum or culture. Furthermore, using just a single topic also means that the focus is on how to teach, rather than what is being taught.
- Longitudinal design: The TALIS Video Study captures student outcome measures before and after they have learnt the focal content, in order to take into consideration students' prior knowledge. Similarly, teachers and students are also surveyed twice to allow for full consideration of their contexts and perceptions.
- Standardised procedures: The TALIS Video Study uses standardised and replicated procedures for data collection, for training and certifying video and artefact raters, and for coding videos and artefacts in every participating school system. This is important because in studies with less stringent processes it can be challenging to determine whether differences across countries are real or simply the result of variation in implementation.

The design features of the observation system

5. The observation codes provide a common international language that makes teaching explicit and facilitates referring to it in a systematic and accurate way across countries. The observation codes specify and delineate the specific constructs and practices of teaching that are measured in the TALIS Video Study.

6. The main goal of the observation codes of the TALIS Video Study is to capture differences in teaching within and between countries. The participating countries have varying levels of student performance, conceptions of teaching quality and cultural backgrounds. Unless these differences are considered, the observation codes would neither measure teaching in a reliable way nor be correlated to other measures of instruction, even if the underlying constructs were actually related.

7. The observation codes are designed to:

- Focus on generic teaching practices: Most of the teaching practices included in the codes can apply to any classroom context, grade and subject. The codes are generic enough to apply across topics in mathematics and can be adapted to be used across secondary school subjects. Therefore, the codes can be a reference to measure teaching in other studies, facilitating systematic comparisons between different instructional contexts and building greater knowledge of how these practices helps students learn and develop.
- Capture a wide range of practices: The codes measure six domains of teaching by looking at 37 different aspects of teaching. The practices that make up the six domains of teaching are broadly defined to avoid promoting any one way of teaching. For example, teachers who score high on 'Aligning instruction to student thinking' might be doing significantly different behaviours, all of which involve adapting their thinking based on student understanding. One teacher might rephrase students' contributions and point out patterns in students' ideas, while another might circulate amongst students, looking at their seatwork, and then reteach an idea with which students are struggling.
- Facilitate a detailed analysis of teaching: The component codes and some indicator codes go beyond describing what happens in the lessons and how frequently it occurs to measure how different teaching practices are performed. For example, the codes differentiate the degree to which the teacher introduces multiple approaches to reasoning rather than if it occurs. In doing so, the observed evidence is placed at the forefront to make the process of analysing teaching transparent and explicit.
- Allow for local application: Given the international scale of the Study, the codes are designed to be applied locally by a significant number of raters in standardised ways. To this end, for example, the training materials for the codes include ground - breaking methods of training adult raters to reason in more standardised ways while still using professional judgement.

8. The codes steer the attention to specific aspects of teaching which are considered of higher quality by the global education community. These aspects of teaching are drawn from the conceptualisation of teaching quality of the TALIS Video Study (see below). Of course there are additional aspects of teaching that are difficult to measure via standardised observation systems (e.g. outside of the classroom practices or teacher self-reflection (Campbell et al., 2003[1]). While some other aspects of teaching are not considered, the Study's codes provide space for holistic professional judgements on more general aspects of classroom interactions such as classroom management or students' cognitive engagement.

The development of the observation system

9. The codes were developed iteratively through four cycles which resulted in five versions across just over two years of development activities. The following goals guided the development process:

Co-designed to build a shared international language

10. Early and regular feedback were considered paramount to ensure that what the codes measure, how constructs or practices are measured, and how ratings are used was appropriate. Participating school systems were fully engaged in the development of the codes, from the very first tasks (defining teaching quality), to the very last tasks (finalising the training and quality control materials). Each development cycle included an initial drafting and testing by the International Consortium, subsequent refining, sharing with experts of the participating countries/economies, and finally revising, retesting and refining again. The Technical Advisory Group also provided feedback to the codes at various stages.

Capacity building for a large scale application

11. To allow for a large scale application, each participating school system led the rating activities under the leadership of the master raters (hereafter referred to as global master raters). The training followed a train-the-trainer model. The International Consortium trained global master raters, who then trained and monitored bilingual raters in their respective school systems using the same materials on which they themselves were trained.

Standardised judgment to ensure comparability

12. The standardisation of the entire observation system was important to reduce the risk of measurement invariance and other validity threats. This is an issue of particular relevance when human raters are involved in scoring (Bell, Carson and Piggott, 2013[2]; Casabianca, Lockwood and McCaffrey, 2015[3]; Floman et al., 2017[4]; Praetorius et al., 2014[5]). Therefore, teaching practices that required raters to have subtle or culturally specific knowledge were either not measured or refined to be understood in a comparable way by all participating school systems. For example, teasing can be a sign of how warmly the teacher feels toward the student in some school systems whilst a sign of disrespect in other ones. The finalised codes did not include teasing as a behaviour, but instead focused on behaviours all systems could agree connoted warmth – e.g. smiling and laughter.

Evidence-centred

13. The codes were developed following an evidence centred design approach (Mislevy, 2011[6]). The scales of each code (e.g. 1-4 or 1-3) were designed to support the following claim: in the observed lesson there was a specific [amount, type, and/or quality] of [teaching practice]. In the development of the codes and training of all raters global master raters were taught to identify and reason with the observable evidence for each practice. This focus on behavioural evidence is the foundation for any subsequent validity claims about teaching practices (Taut, Santelices and Stecher, 2012[7]).

The six domains of teaching measured

14. The conceptualisation of quality teaching resulted from the integration of three bodies of knowledge: participating jurisdictions' views of teaching quality, the global research community's views of good teaching, and the TALIS and PISA specifications of teaching quality.

15. At a high level, these multiple views of teaching quality were aligned however, they regularly emphasised different practices and defined those practices somewhat differently. Further, there were aspects of teaching that were missing from one source or specified in a unique enough manner that they need to be treated as a different aspect of

teaching. In general, these nuances and discrepancies were discussed iteratively with participating school systems until there was a set of constructs that captured teaching in ways that aligned with all three bodies of knowledge.

16. The six domains of teaching measured in the observation codes are classroom management, social-emotional support, discourse, quality of subject matter, student cognitive engagement, and assessment of and responses to student understanding (see below). Each domain is further operationalised into indicators and components depending on whether and how the valued teaching practices can be seen and judged by a rater (see Table 3).

17. The codes capture behaviours that are observable during lessons and about which raters can make inferences without significant additional information from other sources (e.g. an interview with the teacher or the entire quadratic equations unit plan).

Classroom management

18. Classroom management concerns the process of ensuring that lessons run smoothly and efficiently so that teachers' and students' time to focus on academic and social emotional learning is maximised (van Tartwijk and Hammerness, 2011[8]).

19. A particularly important feature of effective classroom management involves establishing and executing routines for common managerial tasks that happen in the classroom on a regular basis (e.g. passing out papers, getting into small groups, beginning the lesson, taking attendance, etc.). Good classroom routines are efficient and help the class to avoid wasting instructional time as much as possible (Anderson, Ryan and Shapiro, 1989[9]; Anderson, Evertson and Emmer, 1980[10]; Muijs and Reynolds, 2000[11]). Effective routines are well organised and carried out smoothly and consistently. They may also support students spending more time on learning activities, thereby maximizing their academic growth (Muijs and Reynolds, 2000[11]).

20. Classroom management also involves the teacher monitoring what is happening across the group of students and proactively addressing issues before they become disruptions. Monitoring can be done in various ways but often features actions such as the teacher maintaining physical proximity to students, scanning the whole classroom from time to time, facing students, calling on a range of students, checking on individual student and group progress, and noticing whether students are on task. In efficient classrooms there is a high ratio of time on task to overall time of the lesson (Prater, 1992[12]).

21. Classrooms are filled with human beings, therefore, disruptions are inevitable. Effective routines and monitoring help reduce the impact of disruptions such as student misbehaviour, external interruptions, or failures of technology when they do occur in the classroom. In a well-managed classroom, these disruptions are addressed quickly and effectively, bringing the instructional activities back on track.

22. Learning takes place in activity structures or what are sometimes referred to as surface features of instruction – whole group, small group, pairs, and individual. While the empirical evidence does not support claims about the general efficacy of certain structures, teachers constantly make decisions about such structures and make use of those activity structures to achieve learning goals. It is important therefore, to understand classroom activity structures within and across jurisdictions.

23. **Social-emotional support**. Group learning of the type students experience in classrooms, requires students to grapple with uncertainty. Such processes require social-emotional support (Klieme, Pauli and Reusser, 2009[13]). An essential element that fosters a supportive learning environment is the positive climate, which is often characterised by

the teacher and students demonstrating respect for one another and regular moments of encouragement and shared warmth in the classroom. In such an environment, respectful language, positive tone of voice and traditional markers of manners are used in verbal communications. Shared warmth such as smiling, laughter, joking, playfulness, enthusiasm or verbal affection are likely to be observed in classroom interactions.

24. Another indicator of social-emotional support involves the degree to which students are willing to take risks in the classroom. When students feel safe, they are more willing to ask questions of and seek guidance from the teacher or other students. They are also more likely to volunteer to share an idea, attempt to articulate an opinion about an issue, or share their private thinking with the whole class (Pianta and Hamre, 2009[14]; Ryan, Gheen and Midgley, 1998[15]; Ryan and Patrick, 2001[16]). Teachers sometimes support this type of social-emotional risk taking by requesting that students share their private thinking with the classroom.

25. Learning necessarily requires that students are intellectually and sometimes emotionally challenged (Ball and Bass, 2000[17]). Such challenge often can be seen behaviourally as errors, misconceptions, or difficulties. It is critical for students to persist through these challenges in order for them to learn (Ball and Bass, 2000[17]; Linnenbrink and Pintrich, 2003[18]). Effective support is demonstrated by teachers and students being patient and encouraging. An acceptance-oriented environment built on trust, should make students feel comfortable and secure, encouraging them to take risks when trying to overcome challenges of various types.

Discourse

26. Classroom discourse – the written and spoken word -- is the medium through which teaching and learning takes place. It is important that there are opportunities for discourse. Students need opportunities to engage in discourse that are clearly focused on a learning objective. It is valuable for students to take a role in such discourse and provide detailed explanations of their thinking so that their thinking becomes visible to peers and the teacher.

27. Discussion, a particular form of discourse, has been documented to be important to student learning (Murphy et al., 2009[19]; Nystrand, 2006[20]). Discussions are extended conversations between and among the teacher and students where students do a good deal of the talking. Although teachers may guide the discussion towards a learning goal, discussions are predominantly based on student ideas and characterised by student-to-student interaction (Franke, Kazemi and Battey, 2007[21]). Discussion opportunities are a potentially important learning opportunity for students (Chapin et al., 2009[22]; Kazemi and Franke, 2004[23]).

28. One major feature of classroom discourse is questioning. Teachers ask many questions in the course of a single lesson (Nystrand et al., 2003[24]; Nystrand et al., 2003[24]). Questioning that facilitates learning requires students to engage in a range of levels of cognitive reasoning that privileges higher order reasoning, which often request students analyse, synthesise, justify, or conjecture (Henningsen and Stein, 1997[25]). Characteristics of such questioning are an appropriate mixture of varied discourse patterns, including IRE (initiate, respond, evaluate), and students speaking back and forth to one another or one after another without the teacher evaluating each student's response. Supportive questioning places the teacher in a facilitating role rather than directing or controlling the discourse without regard for students' contributions (Williams and Baxter, 1996[26]).

29. Another essential element of discourse is explanations (Lachner, Weinhuber and Nückles, 2019[27]). Explanations are descriptions of why ideas or processes are the way

they are (Nunokawa, 2010[28]). In mathematics classrooms, for example, detailed explanations of mathematical ideas or procedures either by the teacher or students support students' learning of mathematics. Well-developed and detailed explanations that focus on deeper features of the mathematics are evidence of thorough understanding of subject matter.

Quality of subject matter

30. While classroom management and social-emotional support will allow classroom interactions to proceed smoothly - with students being on task, engaged and motivated- another important goal of teaching is to promote student interest and understanding of the subject matter. Classrooms that revolve around quality subject matter learning are first and foremost characterised by the clarity and accuracy of the ideas, concepts, and tasks presented. In subject matter rich classrooms, the content in which the teacher and students engage is correct as well as clearly represented so that students are able to focus on understanding the meaning of the concept or task.

31. Student cognition is affected by the explicitness of the learning goals set forth for each lesson (and the extent to which lesson activities are aligned with the learning goal). Student thinking is supported when the teacher clearly communicates the learning goal to students verbally, in written form, or both. Such explicitness supports students' thinking about what they will learn and where it fits with other topics they have learned within that content area or how that idea might connect to their personal experiences or life outside of school.

32. In the case of mathematics, the types of representations are important markers of these subject matter practices. There is not a straightforward relationship that suggests, for example, that as more types of representations are used, students learn more. Quite the opposite might be true. But understanding what type of representations are being used and how they are being used, may lead to new understandings of teaching quality and student learning. In mathematics, for example, multiple representations may be used to support students' understanding.

33. The types and quality of instructional connections can also be indicators of classrooms that are characterised by high levels of subject matter quality. Classrooms are subject matter rich when students and teachers make explicit connections among subject matter ideas, procedures, perspectives, representations or equations that are clear and appropriate. These connections may be experiential connections, where the content being learned is connected to or applied to "real-world" contexts, or subject matter connections, where the content being learned is connected to other topics in the same subject matter or topics in other subject matters (Ball, 1988[29]; Henningsen and Stein, 1997[25]; Leinhardt and Smith, 1985[30]).

34. Explicit patterns and generalizations are important as well. The teacher and students in classrooms with high quality subject matter explicitly look for patterns and generalizations in their work together (Ball, 1988[29]). In addition, they generalise from the content students are working on to a foundational concept and/or definitions underlying the content (Henningsen and Stein, 1997[25]).

35. Quality of subject matter may also be evident in the organization of procedures and content, within and across lessons. The extent to which procedural instructions are clear, correct and well-organised has an impact on whether students are able to make sense of the procedures being taught and apply them appropriately (Ball, 1988[29]). In addition, the quality of lessons is characterised by the presence, clarity, and depth of frequent content summaries where teacher and students explicitly and clearly review or/and summarise what

has been learned (Hospel and Galand, 2016[31]; Kane and Cantrell, 2010[32]; Seidel, Rimmele and Prenzel, 2005[33]). Such summaries can provide students and teachers the opportunity to make sense of the lesson's work or consolidate the knowledge and competencies developed.

Student cognitive engagement

36. As described above, teachers must give students opportunities to engage subject matter practices. But the opportunity to engage in subject matter practices does not necessarily mean that students actually engaged in these practices. Sometimes the teacher engages in the practice, but students only watch. Other times, when students struggle, the teacher changes what she has asked the students to reduce the struggle. So students in the end, do not fully engage in subject matter practices (Baumert et al., 2010[34]; Hiebert and Grouws, 2007[35]; Klieme, Pauli and Reusser, 2009[13]). Again, specific practices depend on the subject matter, but in mathematics classrooms, subject matter practices include engaging in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness (Lipowsky et al., 2009[36]; Mishra and Koehler, 2006[37]; Nunokawa, 2010[28]). The longer and more-often students engage in these practices, the more cognitively active they are likely to be.

37. When students are engaged in cognitively demanding subject matter, in particular, when they work on subject matter procedures and processes, it is important that they use available opportunities to understand why subject matter procedures and processes make sense. For example, students benefit from making sense of individual steps in a mathematical procedure or process; their understanding of the subject matter improves as they attend to the goals and/or properties of procedures and processes, or attend to why a procedure works or a solution is correct (Ball, 1988[29]; Mishra and Koehler, 2006[37]; Nunokawa, 2010[28]).

38. Students' cognitive engagement may be enhanced using multiple approaches to and perspectives on reasoning. For example, in mathematics classrooms, the teacher and students might use two or more procedures or reasoning approaches to solve a problem or type of problem. The depth at which these approaches or perspectives are considered as well as the nature of the similarities and differences across approaches may shape what students learn (Kunter et al., 2013[38]).

39. Attention to metacognition is another critical factor for students' cognitive engagement. A teacher may model self-reflective thinking for students and ask students to reflect on their own thinking in order to develop deeper understandings of their own learning patterns as well as the content and practices being learned (Putnam and Borko, 1997[39]; Schoenfeld, 2016[40]).

40. Learning not only takes place through teacher lecture and modeling, but also through practice opportunities. Practicing may be actions such as writing more than one introductory paragraph for an essay or completing a set of problems with the same underlying theme. Practice opportunities are critical for students to master particular skills through repetition (Ball and Bass, 2000[17]; Ericsson, Krampe and Tesch-Römer, 1993[41]).

41. In a world that is increasingly driven by technology, it is important to understand how it is used in a given discipline and its relation with students' conceptual understanding (Fishman and Dede, 2016[42]). Some technology is used simply to communicate more effectively or efficiently (for example using an overhead projector that allows for coloured markers to highlight key content). Other technology, such as computer programmes or software can be used to plot students' experimental data and quickly calculate equations that describe that data, which can provide students with more robust and evidence-based

opportunities to learn. The type of technology used in the classroom may support certain instructional practices better than others.

Assessment of and Responses to Student Understanding

42. In order to support students' thinking, teachers elicit students' understanding, assess it, and respond to it by aligning their instruction to student thinking. Eliciting student thinking is the first step in this process. In addition, teachers use appropriate questions, prompts, or tasks so that students have opportunities to give answers, but also have the opportunity to explain the reasoning or ideas that supports their answers. A teacher is successful in eliciting student thinking when students' oral and written responses provide detailed evidence of how they understand the process, practices, and ideas pertinent to the subject matter.

43. Once student thinking is elicited, they receive teacher feedback on their thinking. There may be back and forth exchanges, or feedback loops, between the teacher and students that are focused on why the students' understandings are correct or incorrect, and why the ideas and procedures are the way they are. Throughout these exchanges, the teacher uses student responses and actions as a basis for further questioning (Dignath, Buettner and Langfeldt, 2008[43]; Hattie and Timperley, 2007[44]; Kyriakides and Creemers, 2008[45]; Muijs and Reynolds, 2017[46]; Scheerens, 2016[47]). Teachers' feedback may span multiple students' ideas and contributions, eventually leading to an appropriately complete treatment of the subject matter being learnt.

44. In addition to providing feedback to students, teachers use students' responses productively by aligning instruction to present student understanding. Teachers may align instruction in a variety of ways. For example, they might notice student thinking by simply circulating in the classroom and looking at students' papers; they may review homework problems and notice common patterns that shape subsequent actions, or they may comment to students about the understandings in the classroom. There are diverse ways that teachers use students' contributions and support student understanding (Borko and Livingston, 1989[48]). For example, if students have a misconception, make an error, state an observation, or ask a question, the teacher will use those responses to help improve students' understanding. The teacher many draw attention to the contribution, ask a question in response to a student's question, have students provide the next step in the procedure, or acknowledge patterns in student contributions. A key marker of instructional alignment is that the teacher provides hints and cues when students struggle mathematically or make errors (Hayes, 2003[49]; Taylor, 2007[50]).

Table 1. TALIS-Video Domains, Holistic domain ratings, Components, and Indicators

Domain	Components	Indicators
Classroom management	Routines	Time on task
	Monitoring	Activity structure and frequency
	Disruptions	Time of lesson (only after last segment)
	Classroom management (overall)	
Social-emotional support	Respect	
	Encouragement and warmth	Persistence
	Risk-taking	Requests for public sharing
	Social-emotional support (overall)	
Discourse	Nature of discourse	
	Questioning	Discussion opportunities
	Explanations	
	Discourse (overall)	
Quality of subject matter		Explicit learning goals
		Accuracy
	Explicit connections	Real-world connections
	Explicit patterns and generalisations	Connecting mathematical topics
	Clarity	Mathematical summary
	Quality of subject matter (overall)	Types of representation
		Organisation of procedural instruction
Student cognitive engagement		Metacognition
	Engagement in cognitively demanding subject matter	Repetitive use opportunities
	Multiple approaches to/perspectives on reasoning	Technology for understanding
	Understanding of subject matter procedures and processes	Classroom technology
	Student cognitive engagement (overall)	Student technology
		Software use for learning
Assessment of and responses to student understanding	Eliciting student thinking	
	Teacher feedback	
	Aligning instruction to present student thinking	
	Assessment of and responses to student understanding (overall)	

The TALIS Video observation system

45. Whenever an observation rubric or checklist is applied to a lesson, it is applied in an observation system (Hill, Charalambous and Kraft, 2012[51]). Observation systems are comprised of rating specifications, rating processes, and sampling and scoring specifications (Bell et al., 2018[52]; Liu et al., 2019[53]) (Table 2). The extent to which each of them are specified varies across observation systems. The approach of the TALIS Video Study to produce valid and reliable scores is described below.

Table 2. Observation system components

Rating Specifications	Rating Processes	Sampling and Scoring
Aspects of teaching	Training	Student sampling
Scales	Certification	Time sampling
Standards in master ratings	Calibration	Subject matter sampling
	Validation	Scoring model
	Multiple ratings	
	Rater assignment	

Source: (Liu et al., 2019[53])

Rating specifications

46. Aspects of teaching, scales, and standards in master ratings are the tools raters use to assign ratings; they are scoring tools. Aspects of teaching are the dimensions that codify teaching – e.g. eliciting student thinking, persistence, accuracy. In the TALIS Video Study, the six domains of teaching include 18 component codes and 19 indicator codes (Table 1). The component and indicator definitions specify the particular aspects of teaching measured in the Video Study.

47. Scales are the categories raters use to describe the nature, quality, presence, or frequency. Components are rated on a scale from one to four, whereas most indicators are rated on a scale from one to three or are descriptive noting the presence or absence of a specific classroom structure (e.g. the use of small groups, the presence of a graphical representation the use of software). Together, the specific aspects of teaching and their scales are referred to as codes. The codes of the TALIS Video Study are shown in Annex B.

48. Using video examples of what counts in the codes, the master ratings set standards to move from behaviours in a classroom to words on a page to the numerical rating that can be aggregated and analysed. Global master raters, who have a deep understanding of the observation codes, create the "gold standard" that raters use as benchmarks to guide their application of codes' words to lesson behaviours. Master rated videos that show, for example, what a two versus as three looks like for the routines component, help raters understand how to apply the scales accurately.

Rating processes

49. The rating processes aim to ensure that raters are able to reliably and accurately apply the rating specifications. The TALIS Video Study observation system includes the following rating process components:

Training

50. The goal of the training was to help raters understand the aspects of teaching measured in the system as well as the rating scales. Global master raters were trained face-to-face for three days on the components codes and two days on the indicator codes. Training was structured as a whole group activity in which global master raters had computers that they used for practicing rating and accessing training materials. All training materials were in English or had English subtitles (in the case of the videos). Global master raters were expected to use the exact same approach and materials to train school system raters.

51. A rather novel approach of the study is the focus on how raters took notes and how they were expected to think. Building upon research on rater thinking and performance, the rubrics and training materials were adjusted on the basis of shared understanding built by raters contributions and repeated opportunities for raters to practice using the codes. Raters were asked to identify the specific behaviours that might count as evidence for a particular code and conversations followed to standardise how raters' thought with the evidence and the rating scales. The training materials provided guiding questions to discipline raters' thinking, specified rules for common difficulties such as how to think about contradictory evidence, provided guidance about where and how to count specific evidence in the rating scales, and provided examples of common behaviours that should be applied to various codes. The training approaches taken were similarly effective across jurisdictions (see Annex C).

Certification

52. The certification process tested whether raters had acquired in the training the knowledge and skills required to apply the rating specifications. At the end of each of the training sessions for the component and indicator codes, global master raters took a certification test comprised of two lessons. Each lesson was from a different participating school system, and varied in length but was no shorter than 32 minutes. Table 3 shows the raters minimum level of agreement required to pass the certification and average level of agreement reached on the certification test for raters in the participating school systems. All global master raters passed on their first attempt.

Table 3. Raters levels of agreement with the standards (percentage)

	Indicators		Components	
Level of agreement	Minimum required	Average	Minimum required	Average
Exact	75	84	50	60
Exact and adjacent	90	96	85	96

53. Raters had to meet the same standards as the global master raters. For indicators, 93% of raters trained in the eight jurisdictions passed the certification test on their first attempt. For components, 98% passed on their first certification attempt. The number of raters certified varied across participating school systems from a low of seven in Japan to a high of twenty-six in Colombia. Annex C provides detailed jurisdiction rater to rater agreement metrics, by domain for both indicators and components. As has been documented elsewhere (Praetorius et al., 2014[5]; Bell et al., 2015[54]) agreement rates were highest on the classroom management codes and lowest on the instructionally focused codes.

Calibration

54. Calibration events provided raters with opportunities to get feedback on their ratings, and continue to learn and calibrate their application of the rating specifications. After main study rating began, raters participated in up to eight regular calibration events. They were asked to rate 1-2 segments of a lesson from one of the eight school systems which had already been master scored by the global master raters. Then, global master raters discussed ratings, discrepancies in their understanding of codes, and challenges or concerns of the scoring process with the jurisdiction raters.

Validation

55. Validation videos measure the degree of adherence to the rating specifications throughout the rating period. Raters rated up to six validation videos, roughly after every six videos, without being told that they were being monitored. The global master raters also scored these videos. Results for both calibration and validation videos in each jurisdiction will be analysed to evaluate accuracy and reliability and report. High level results of certification, calibration and validation processes for components and indicators are in Annex C. More detail will be available in the TALIS Video study technical report.

Other rating processes

56. Other rating processes used in the TALIS Video Study include:

- Double rating: Each lesson is rated by two independent raters. The multiple ratings collected are used to calculate metrics of inter-rater agreement (see Annex C for results).
- Random rater assignment. Careful rater assignment can lead to higher quality scores. Rater assignment is the process used to assign raters to teachers, schools, or classrooms so that no one rater has an undue impact on the ratings for that unit of analysis. For example, if four videos from a single teacher must be rated, raters would be assigned to those four videos so that the largest number of different raters could rate those four videos.

57. Although insights regarding the validity of the TALIS Video codes will not be available until 2020 after the data is analysed, the preliminary training and certification evidence suggests the approaches used for code development and training are scalable.

Sampling and scoring

58. Observation systems delineate sampling and scoring specifications in order to determine what is to be analysed. The section on ‘The design of the study’ provides greater information on the sampling of students, time and subject matter in the TALIS Video Study. The entire length of the video recorded for the two lessons was scored. Raters were instructed to divide lessons in 16 minutes components and 8 minutes indicators for the scoring as explained in Annex B.

References

Anderson, L., C. Evertson and E. Emmer (1980), "Dimensions in classroom management derived from recent research", *Journal of Curriculum Studies*, Vol. 12/4, pp. 343-356, http://dx.doi.org/10.1080/0022027800120407. [10]

Anderson, L., D. Ryan and B. Shapiro (1989), *The IEA Classroom Environment Study*, Pergamon Press, PLC. [9]

Ball, D. (1988), "Research on teaching mathematics: Making subject matter knowledge part of the equation", Office of Educational Research and Improvement, Washington, DC. [29]

Ball, D. and H. Bass (2000), "Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics", in Boaler, J. (ed.), *Multiple Perspectives on the Teaching and Learning of Mathematics*, Ablex Publishing, Westport, NY. [17]

Baumert, J. et al. (2010), "Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress", *American Educational Research Journal*, Vol. 47/1, pp. 133-180, http://dx.doi.org/10.3102/0002831209345157. [34]

Bell, A., L. Carson and L. Piggott (2013), "Deliberative democracy for curriculum renewal", in Dunne, E. and D. Owen (eds.), *The student engagement handbook: Practice in higher education*, Emerald, Bingley, UK. [2]

Bell, C. et al. (2018), "Qualities of classroom observation systems", *School Effectiveness and School Improvement*, pp. 1-27, http://dx.doi.org/10.1080/09243453.2018.1539014. [52]

Bell, C. et al. (2015), "Improving observational score quality: Challenges in observer thinking", in Kane, T., A. Kerr and K. Pianta (eds.), *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, John Wiley & Sons, San Francisco, CA., http://dx.doi.org/10.1002/9781119210856.ch3. [54]

Borko, H. and C. Livingston (1989), "Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers", *American Educational Research Journal*, Vol. 26/4, pp. 473-498, http://dx.doi.org/10.3102/00028312026004473. [48]

Campbell, R. et al. (2003), "Differential teacher effectiveness: Towards a model for research and teacher appraisal", *Oxford Review of Education*, Vol. 29/3, pp. 347-362. [1]

Casabianca, J., J. Lockwood and D. McCaffrey (2015), "Trends in classroom observation scores", *Educational and Psychological Measurement*, Vol. 75/2, pp. 311-337, http://dx.doi.org/10.1177/0013164414539163. [3]

Chapin, S. et al. (2009), *Classroom discussions: Using math talk to help students learn, Grades K-6*, https://books.google.com/books?hl=en&lr=&id=2NX4I6mekq8C&oi=fnd&pg=PR15&dq=Chapin,+S.+H.,+C.+O%E2%80%99Connor+and+N.+C.+Anderson+(2009),+Classroom+Discussions:+Using+Math+Talk+to+Help+Students+Learn,+Grades+K-6,+Math+Solutions,+Boston.+&ots=4Y2I-Armz0&sig=8fC8CVUSzekIMlQM4JNYpzJMle4 (accessed on 16 December 2019). [22]

Dignath, C., G. Buettner and H. Langfeldt (2008), *How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes*, Educational Research Review, http://dx.doi.org/10.1016/j.edurev.2008.02.003. [43]

Ericsson, K., R. Krampe and C. Tesch-Römer (1993), "The role of deliberate practice in the acquisition of expert performance", *Psychological Review*, Vol. 100/3, pp. 363-406, http://dx.doi.org/10.1037/0033-295x.100.3.363. [41]

Fishman, B. and C. Dede (2016), "Teaching and Technology: New Tools for New Times", in *Handbook of Research on Teaching*, http://dx.doi.org/10.3102/978-0-935302-48-6_21. [42]

Floman, J. et al. (2017), "Emotional Bias in Classroom Observations: Within-Rater Positive Emotion Predicts Favorable Assessments of Classroom Quality", *Journal of Psychoeducational Assessment*, Vol. 35/3, pp. 291–301, https://doi.org/10.1177/0734282916629595. [4]

Franke, M., E. Kazemi and D. Battey (2007), "Mathematics teaching and classroom practice", in Lester, F. (ed.), *Second Handbook of Research on Mathematics Teaching and Learning*, Information Age Publishing, Charlotte, NC., https://www.researchgate.net/publication/285328217 (accessed on 17 December 2019). [21]

Hattie, J. and H. Timperley (2007), "The Power of Feedback", *Review of Educational Research*, Vol. 77/1, pp. 81-112, http://dx.doi.org/10.3102/003465430298487. [44]

Hayes, D. (2003), "Making learning an effect of schooling: Aligning curriculum, assessment and pedagogy", *Discourse*, Vol. 24/2, pp. 225-245, http://dx.doi.org/10.1080/01596300303039. [49]

Henningsen, M. and M. Stein (1997), "Mathematical Tasks and Student Cognition: Classroom-Based Factors That Support and Inhibit High-Level Mathematical Thinking and Reasoning", *Journal for Research in Mathematics Education*, Vol. 28/5, p. 524, http://dx.doi.org/10.2307/749690. [25]

Hiebert, J. and D. Grouws (2007), "The effects of classroom mathematics teaching on students' learning", in Leste, F. (ed.), *Second Handbook of Research on Mathematics Teaching and Learning*, Information Age, Charlotte, NC. [35]

Hill, H., C. Charalambous and M. Kraft (2012), "When rater reliability is not enough", *Educational Researcher*, Vol. 41/2, pp. 56-64, http://dx.doi.org/10.3102/0013189X12437203. [51]

Hospel, V. and B. Galand (2016), "Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis", *Learning and Instruction*. [31]

Kane, T. and S. Cantrell (2010), "Learning about teaching initial findings from the measures of rffective teaching project about the measures of effective teaching project", in *MET Project Research Paper*, Bill & Melinda Gates Foundation. [32]

Kaput, J. (1999), "Teaching and learning a new algebra", in Romberg, T. and E. Fennema (eds.), *Mathematics classroom that promotes understanding*, Lawrence Erlbaum, Hillsdale. [55]

Kazemi, E. and M. Franke (2004), "Teacher learning in mathematics: Using student work to promote collective inquiry", *Journal of Mathematics Teacher Education*, Vol. 7/3, pp. 203-235, http://dx.doi.org/10.1023/B:JMTE.0000033084.26326.19. [23]

Klieme, E., C. Pauli and K. Reusser (2009), "The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms", in Janik, T. (ed.), *The Power of* [13]

Video Studies in Investigating Teaching and Learning in the Classroom, Waxmann, Münster, Germany, https://www.recherche-portal.ch/zbz/action/display.do?fn=display&vid=ZAD&doc=ebi01_prod006351395.

Kunter, M. et al. (eds.) (2013), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*, Springer, New York. [38]

Kyriakides, L. and B. Creemers (2008), "Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model", *School Effectiveness and School Improvement*, Vol. 19/2, pp. 183-205, http://dx.doi.org/10.1080/09243450802047873. [45]

Lachner, A., M. Weinhuber and M. Nückles (2019), "To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of Principle-Oriented explanations", *Contemporary Educational Psychology*, Vol. 58, pp. 175-185, http://dx.doi.org/10.1016/j.cedpsych.2019.03.008. [27]

Leinhardt, G. and D. Smith (1985), "Expertise in mathematics instruction. Subject matter knowledge", *Journal of Educational Psychology*, Vol. 77/3, pp. 247-271, http://dx.doi.org/10.1037/0022-0663.77.3.247. [30]

Linnenbrink, E. and P. Pintrich (2003), "The role of self-efficacy beliefs instudent engagement and learning intheclassroom", *Reading and Writing Quarterly*, Vol. 19/2, pp. 119-137, http://dx.doi.org/10.1080/10573560308223. [18]

Lipowsky, F. et al. (2009), "Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem", *Learning and Instruction*, Vol. 19/6, pp. 527-537, http://dx.doi.org/10.1016/j.learninstruc.2008.11.001. [36]

Liu, S. et al. (2019), "Classroom observation systems in context: A case for the validation of observation systems", *Educational Assessment, Evaluation and Accountability*, Vol. 31/1, pp. 61-95, http://dx.doi.org/10.1007/s11092-018-09291-3. [53]

Mishra, P. and M. Koehler (2006), "Technological pedagogical content knowledge: A framework for teacher knowledge", *Teachers College Record*, Vol. 108/6, pp. 1017-1054, http://dx.doi.org/10.1111/j.1467-9620.2006.00684.x. [37]

Mislevy, R. (2011), "Evidence-Centered Design for Simulation-Based Assessment", No. 800, The National Center for Research on Evaluation, Standards, and Student Testing, University of California at Los Angeles. [6]

Muijs, D. and D. Reynolds (2017), *Effective teaching: evidence and practice*, SAGE Publications Ltd. [46]

Muijs, D. and D. Reynolds (2000), "School effectiveness and teacher effectiveness in mathematics: Some Preliminary findings from the evaluation of the mathematics enhancement programme (primary)", *School Effectiveness and School Improvement*, Vol. 11/3, pp. 273-303, http://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273. [11]

Murphy, P. et al. (2009), "Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis", *Journal of Educational Psychology*, Vol. 101/3, pp. 740-764, http://dx.doi.org/10.1037/a0015576. [19]

Nunokawa, K. (2010), "Proof, mathematical problem-solving, and explanation in mathematics [28] teaching", in G., H., J. H. and P. H. (eds.), *Explanation and Proof in Mathematics*, Springer, Boston, MA, http://dx.doi.org/10.1007/978-1-4419-0576-5_15.

Nystrand, M. (2006), "Research on the role of classroom discourse as it affects reading comprehension", *Research in the Teaching of English*, Vol. 40, pp. 392-412, http://www.jstor.org/stable/401717. [20]

Nystrand, M. et al. (2003), "Questions in time: Investigating the structure and dynamics of unfolding classroom discourse", *Discourse Processes*, Vol. 35/2, pp. 135-198, http://dx.doi.org/10.1207/s15326950dp3502_3. [24]

Pianta, R. and B. Hamre (2009), "Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity", *Educational Researcher*, Vol. 38/2, http://dx.doi.org/10.3102/0013189X09332374. [14]

Praetorius, A. et al. (2014), "One lesson is all you need? Stability of instructional quality across lessons", *Learning and*, https://www.sciencedirect.com/science/article/pii/S0959475213000832 (accessed on 3 March 2020). [5]

Prater, M. (1992), *Increasing Time-on-Task in the Classroom Suggestions for improving the amount of time learners spend in on-task behaviors*. [12]

Putnam, R. and H. Borko (1997), "Teacher learning: Implications of new views of cognition", in Biddle, B., T. Good and I. Goodson (eds.), *International Handbook of Teachers and Teaching Vol.3.*, Springer, Dordrecht, http://dx.doi.org/10.1007/978-94-011-4942-6_30. [39]

Ryan, A., M. Gheen and C. Midgley (1998), "Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social–emotional role, and the classroom goal structure", *Journal of Educational Psychology*, Vol. 90/3, pp. 528-535, http://dx.doi.org/10.1037/0022-0663.90.3.528. [15]

Ryan, A. and H. Patrick (2001), "The classroom social environment and changes in adolescents' motivation and engagement during middle school", *American Educational Research Journal*, Vol. 38/2, pp. 437-460, http://dx.doi.org/10.3102/00028312038002437. [16]

Scheerens, J. (2016), "Meta-analyses of school and instructional effectiveness", in Scheerens, J. (ed.), *Educational Effectiveness and Ineffectiveness*, Springer Netherlands, http://dx.doi.org/10.1007/978-94-017-7459-8_8. [47]

Schoenfeld, A. (2016), "Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics", *Journal of Education*, Vol. 196/2, pp. 1-38, http://dx.doi.org/10.1177/002205741619600202. [40]

Seidel, T., R. Rimmele and M. Prenzel (2005), "Clarity and coherence of lesson goals as a scaffold for student learning", *Learning and Instruction*, Vol. 15/6, pp. 539-556, http://dx.doi.org/10.1016/j.learninstruc.2005.08.004. [33]

Taut, S., M. Santelices and B. Stecher (2012), "Validation of a national teacher assessment and improvement system", *Educational Assessment*, Vol. 17/4, pp. 163-199. [7]

Taylor, J. (2007), *Alignment of Teachers' Educational Goals and Habits of Understanding, Action, Reflection, and Improving Practice with the Iowa State University Secondary Science Teacher Education Program*, Iowa State University, Ames, http://dx.doi.org/10.31274/rtd-180813- [50]

16233.

van Tartwijk, J. and K. Hammerness (2011), “The neglected role of classroom management in teacher education”, *Teaching Education*, Vol. 22/2, pp. 109-112, http://dx.doi.org/10.1080/10476210.2011.567836. [8]

Williams, S. and J. Baxter (1996), “Dilemmas of Discourse-Oriented Teaching in One Middle School Mathematics Classroom”, *The Elementary School Journal*, Vol. 97/1, pp. 21-38. [26]

Annex A. The procedures for rating

Annex A contains the training manual developed for raters of the TALIS Video Study. It explains the procedures to be followed in the rating of indicators and components.

Rating terminology

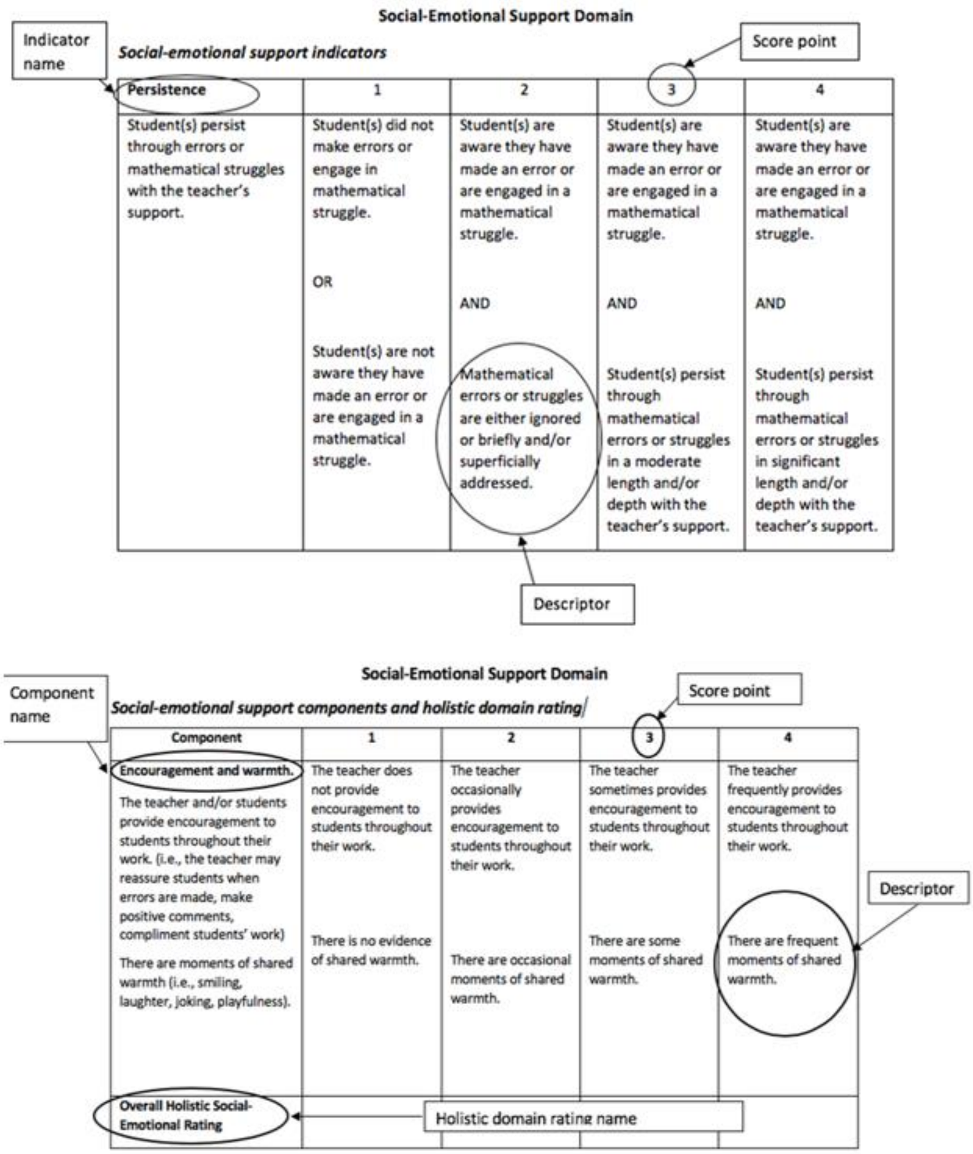
59. The resources developed to help raters code the TALIS Video lessons were: the training manual, the training slides, and the training videos. The codes provide the rating scales, the training manual and slides explain how to apply those scales and provides critical definitions and examples of the scales, and the training videos provide embodied examples of the application of the codes to actual classroom interactions. Rating videos of classroom interactions is complex work and requires a strong understanding of the codes.

60. There are a number of key terms used in the training manual. They are defined below and represented in Figure A.1.

Table A A.1. Important TALIS Video Study terms and definitions

Term	Definition
Domain	One of the six aspects of teaching that supports students' learning and is measured by the TALIS Video observation codes. Each domain is comprised of components (which the rater uses to assign holistic domain ratings) and indicators.
Component	A code that applies to higher inference classroom interactions and rates the interactions on a four-point scale. They are rated every 16 minutes.
Holistic domain rating	A holistic code that is created by the rater after rating the components. Due to the dependence of holistic domain ratings on components, they necessarily apply to higher inference classroom interactions and are rated on a four-point scale.
Indicator	A code that applies to lower inference classroom interactions and either categorises or rates the interactions on a scoring scale. They are rated every 8 minutes.
Descriptor	A description of the interactions that characterise each score point of components or indicators.
Score point	The numerical rating assigned by raters to each component, holistic domain rating, and indicator based on the associated descriptor.

Figure A A.1. Key terms



Persistence	1	2	3	4
Student(s) persist through errors or mathematical struggles with the teacher's support.	Student(s) did not make errors or engage in mathematical struggle. OR Student(s) are not aware they have made an error or are engaged in a mathematical struggle.	Student(s) are aware they have made an error or are engaged in a mathematical struggle. AND Mathematical errors or struggles are either ignored or briefly and/or superficially addressed.	Student(s) are aware they have made an error or are engaged in a mathematical struggle. AND Student(s) persist through mathematical errors or struggles in a moderate length and/or depth with the teacher's support.	Student(s) are aware they have made an error or are engaged in a mathematical struggle. AND Student(s) persist through mathematical errors or struggles in significant length and/or depth with the teacher's support.

Component	1	2	3	4
Encouragement and warmth. The teacher and/or students provide encouragement to students throughout their work. (i.e., the teacher may reassure students when errors are made, make positive comments, compliment students' work) There are moments of shared warmth (i.e., smiling, laughter, joking, playfulness).	The teacher does not provide encouragement to students throughout their work. There is no evidence of shared warmth.	The teacher occasionally provides encouragement to students throughout their work. There are occasional moments of shared warmth.	The teacher sometimes provides encouragement to students throughout their work. There are some moments of shared warmth.	The teacher frequently provides encouragement to students throughout their work. There are frequent moments of shared warmth.
Overall Holistic Social-Emotional Rating				

The rating process

61. Indicators are rated every 8 minutes, while components and holistic domain ratings are rated at 16-minute segments throughout the lesson. Each segment should be treated independently. Exceptions to this are noted in the manual.

Transcripts

62. Many countries rate using transcripts. These countries will mark up the transcripts as directed, using shorthand and highlighting tools. A few countries will not work with transcripts and instead plan to take notes while the video is playing. For these countries, raters must take notes in their native language for the entire segment. To the degree possible, the raters should use abbreviations and shorthand in order to scribe what is being said by the students and teacher. The goal will be to take unbiased notes that capture the

conversation between and among teachers and students. These notes will be collected as part of the main study data set.

Pace

63. Raters may start and stop the videos as needed to take notes and/or annotate the transcript. At the end of the segment, raters will pause the video and then assign ratings for indicators or for all of the holistic domain ratings and components.

Scale

64. Raters rating the indicators will use the varied scales and categories required by the indicators, most commonly using a 3-point scale.

65. Raters will first record their component ratings on a 4-point scale and then assign a holistic domain rating on the same 4-point scale. Overall holistic domain rating ratings are holistic and should not be thought of as a mathematical “average” of the components. However, raters should only use information the components ask them to attend to. They should not introduce new constructs into their holistic rating. Instead, they should consider the nature and weight of the evidence across all components when assigning the holistic domain rating. Only the analysis of the data will determine which set of ratings (i.e. aggregated components or rater assigned holistic domain ratings) is more reliable and associated with student outcomes.

The cognitive process of rating

66. The rater should always begin by reading the description of the indicator or component in the far left box of the rubric and refer to it frequently. [S]he should then go to the lowest level on the scale (the 1 score point for indicators/components or the lowest level) and read the first descriptor. The rater should then review her notes for evidence that would support that descriptor. If the evidence does not support that descriptor, the rater should move up to the next score point. This process should continue until the rater finds the proper score point for the descriptor and evidence. If there is more than one descriptor for a given score point (see Figure A.1), the rater should carry out the process multiple times. This may lead to different score points for descriptors within a given indicator or component. If this occurs, the rater must make a judgment about which rating to assign based on the preponderance of evidence (explained further below), the benchmarks, and the rater’s understanding of the distribution of videos across all training videos. For indicators that are categorical, the rater should read through all categories every time that indicator is rated.

67. Some of the teaching practices described in the protocol are expected to be rare, and therefore they may appear infrequently or not at all in some lessons. Also, the teaching practices described in the protocol are unlikely to appear all at once in one segment of instruction or even in one lesson. For example, a lesson with clear instructional content being delivered might not present opportunities for the teacher and students to show persistence through challenges. Giving this video a 1 in Persistence does not mean the teacher is doing a poor job. Rather, it means that no challenges arose during the segment of instruction being coded. This example is applicable to other codes in the protocol as well (e.g. mathematical summary, real-world connections). Some of the indicator codes, are more descriptive in nature (e.g. time of the lesson, activity structure, technology) and should not be thought about in terms of quality.

68. The goal of the TALIS Video Study is to understand the range of teaching practices that appear in classrooms around the world. The first step is to accurately record what is happening, the analysis will determine within each country, which practices are associated with positive student outcomes. The practices related to positive outcomes in one country might not be positively related to outcomes in other countries (e.g. working regularly in small groups, using technology, etc.). Your goal is to accurately document what occurs in the lessons so that the analysis can determine how practices are related to outcomes across the lessons in your country's study.

Assumptions about teacher's intent

69. Because of how the TALIS Video Study samples instruction – i.e. teachers are asked to allow the recording of lessons on quadratic equations – the rater should assume that the activities in the recorded lessons are related to quadratic equations, unless there is very strong evidence to the contrary.

Consulting artefacts

70. Many times teachers repeat or read aloud whatever is written on the board or in front of students on their desks. In these cases, the rater does not have to review the artefacts. If the teacher does not do this, the rater should always review the artefacts to assist with the rating process. When students work independently at their desks, the rater should always know what problems are being worked on. This is especially important in making judgments about student cognitive engagement, quality of subject matter, and discourse (because written discourse must be counted where possible).

Preponderance of evidence

71. It is highly likely that evidence will not line up neatly under a single score point. The rater must then make a judgment about which score point is best supported by the evidence. For example, there may be some evidence in a segment that falls under a rating of 2, but if most of the evidence falls under a rating of 3, then a rating of 3 may be more applicable. In the specific case of components that have more than one descriptor, the evidence might support two different score points and the rater will have to make a judgment about which score point is most able to be supported by the evidence. The same is true for holistic domain ratings. The rater will need to consider the evidence across the components, which often will vary across score points. For example, if one component rates a 4, but the other two components rate a 1, the rater must make the decision to rate the performance somewhere in between the score points based on the preponderance of evidence.

72. The process raters should follow when deciding on the preponderance of evidence should be guided by the following examples:

Holistic ratings

73. When deciding between a holistic rating of 2 or 3 the rater should ask three questions to help decide:

1. What is the strength of the evidence I have for each component?
2. Did the evidence in this segment fall in the top half or bottom half of the distribution of segments I have seen for this code?

3. (Metacognitive check) Am I introducing criteria other than the component criteria in deciding on the score?

74. When deciding between a 3 and a 4 remember that the segment does not need to be "perfect" on all three components to be rated a 4, i.e. all 4s on the three components. However, there should be consistent evidence that goes across the segment and across all three components that comprise the domain.

Component ratings:

75. When deciding how to rate a component whose descriptors split across two score points (e.g. 1 and 4 or 2 and 3), the rater should ask three questions to help decide:

1. What is the strength of the evidence I have for each descriptor?
2. Did the evidence in this segment fall in the top half or bottom half of the distribution of segments I have seen for this code?
3. (Metacognitive check) Am I introducing criteria other than the component criteria in deciding on the score?

Behaviours counting for more than one component/holistic domain rating

76. Teaching and learning are complex interactional processes that evolve over time with multiple individuals. A single action or interaction might mean different things, depending on which aspect of teaching is being judged. For example, the types of questions a teacher poses may be considered evidence to support ratings in the Discourse domain as well as the Student Cognitive Engagement domain. The rater must therefore, examine the criteria in each domain as well as the evidence in each performance descriptor when rating.

Whose behaviour counts?

77. Each component/holistic domain rating or indicator requires the rater to pay attention to specific evidence in order to rate. That evidence comes from one of three places: the teacher, the students, or both the teacher and the students. Written behaviour also should be counted. For each component/holistic domain rating, the source of evidence is noted.

Judging the frequency of certain behaviours

78. Often the score descriptor will have a word that refers to the frequency with which some behaviour occurred. That frequency should be judged against the segment length (e.g. 8 or 16 minutes). It also should take account of the rater's judgment. The following guidelines should guide raters to their ratings; however, the rater should not use the rules below formulaically. Further, benchmarks of various score points shown during training should also be used to guide raters. The rater should remember that the benchmarks are normed against the global distribution of teaching and that distribution is especially important to keep in mind. Table A.2 should be read as guidance, not a rule.

Table A A.2. Frequency language and number of occurrences

Frequency	Components	Indicator
Rare / occasionally	1-2	1
A handful / a couple / few	2-3	2
Some / sometimes / often / most	3-4	3
Frequent	4+	4
All	All	All

Rating the last segment

79. At the end of a video when a segment is less than 8 minutes for indicators and 16 minutes for components, raters will carry out one of two actions depending on the remaining minutes:

1. If greater than or equal to 4 minutes for indicators and 8 minutes for components, create a new segment.
2. If less than 4 minutes for indicators and 8 minutes for components, append the remaining minutes to the previous segment.

Table A A.3. Exemplar video lengths and number of component and indicator segments

Video length	Number of holistic domain ratings / component segments	Video length	Number of indicator segments
32:00	2	32:00	4
36:03	2	36:03	5
40:00	3	40:00	5
42:15	3	42:15	5
48:00	3	48:00	6

Weighting of component descriptors within a component

80. Some component score points have more than one descriptor (e.g. see Figure A.1). The descriptors are <u>not</u> listed in order of importance. They are equally important. Raters should rate each descriptor separately. If there are differences (e.g. one descriptor has evidence to support a 2 and the second descriptor has evidence to support a 4), the rater should weigh all the descriptors equally and evidence holistically and determine an overall rating. Usually this will lead to a rating in the middle of the two discrepant descriptors. Benchmarks can assist in making decisions.

Annex B. The Observation Codes

Annex B presents an abridged version of the observation code. The versions used by raters will be published verbatim in the technical report of the TALIS Video Study.

Classroom Management

Table A B.1. Routines

Component	1	2	3	4
Routines. The classroom has routines for common managerial tasks that are organised and efficient.	A small proportion of routines are organised. Routines frequently waste time.	A moderate proportion of routines are organised. Routines sometimes waste time.	A large proportion of routines are organised. Routines rarely waste time.	All routines are organised. Routines do not waste time.

Definition of routines

81. Routines are common tasks, chores, or duties that should be done regularly or at specified intervals typically, an everyday activity.

Notes and rating guidance

- The mathematics being worked on should not be the focus when the rater is rating routines.
- There is almost always at least one routine to consider.
- Examples of routines include but are not limited to: passing out papers, getting into small groups, beginning the lesson, small groups reporting out, getting the teacher's attention during individual seatwork, procedures for sharing student ideas in a whole group format, greeting one another at the beginning of a lesson, taking attendance, taking notes when the teacher is writing on the board, and checking homework.
- Raters should ask first: What routines did I see? If there is no evidence of any routines to rate, the component should be scored as a 1.
- Whenever students are carrying out tasks that are unrelated to the establishment or execution of regular routines (e.g. they get up to leave for a fire drill or they get interrupted by the loudspeaker) this should not be considered in rating routines. The rater should account for these inefficiencies in disruptions.
- If there is wasted time in a routine, the rater should not assign a 4.
- Evidence: teacher and students.

Table A B.2. Monitoring

Component	1	2	3	4
Monitoring. The teacher monitors what is happening in the entire classroom (i.e., the teacher maintains physical proximity to students, scans the whole classroom, faces students, calls on a range of students, and notices student behavioural progress).	There is little or no evidence that the teacher monitors what is happening in the entire classroom.	The teacher occasionally monitors the entire classroom.	The teacher sometimes monitors the entire classroom and monitoring may have inconsistencies.	The teacher frequently monitors the entire classroom and does so consistently.

Definition of monitoring

82. A teacher maintains physical proximity to students, scans the entire classroom, faces students, calls on a range of students and notices student progress while monitoring the classroom.

Notes and rating guidance

- A teacher does not need to do all of these monitoring behaviours in order for the segment to be rated a 4.
- The component focuses on monitoring students' progress through routines, tasks, and activities of the classroom. The rater should account for the degree to which the teacher monitors students' behavioural participation and progress, not on students' intellectual progress.
- Whatever monitoring behaviours a teacher uses should be judged on the degree to which she or he uses those strategies for the whole class, not just a small group of students.
- Rating guidance
    - If there is off-task behaviour that the teacher does not monitor, the rater generally should not assign a 4 rating.
    - Frequently, students are quiet and attentive and teachers do not carry out a lot of monitoring actions. Such a segment would be rated lower than a segment in which the students behave in a similar fashion and there are more monitoring behaviours.
    - If the teacher only monitors one side of the classroom, the rater should generally not assign a 4 rating.
- Evidence: teacher and students.

Table A B.3. Disruptions

Component	1	2	3	4
Disruptions. The teacher quickly and effectively deals with disruptions. There are few or no disruptions.	The teacher does not handle disruptions effectively or efficiently, causing the class to lose significant instructional time.	The teacher may occasionally handle disruptions effectively, but in general, the teacher does not effectively or efficiently handle disruptions, causing the class to lose some instructional time.	The teacher generally handles disruptions effectively, but sometimes disruptions cause the class to lose a small amount of instructional time.	The teacher handles disruptions quickly and effectively so that instructional time may be interrupted but not lost. OR There are no disruptions

Definition of disruptions

83. A disruption is an instance when teachers', students' or external actors' behaviour(s) draws significant attention away from the subject matter or classroom activities.

Notes and rating guidance

- Student behaviours that cause disorder or turmoil to the lesson should be accounted for as a disruption.
- To understand whether a disruption is occurring in a specific culture, the rater must attend to how the other students and teacher react to the behaviour. A student eating food in class might not be a disruption in a classroom in one country's context but in another, it is a disruption.
- Examples of disruption: student misbehaviour, external interruptions, off-topic loudness and noise, failures of technology, etc.
- Rating guidance:
  - If no disruptions arise and it is clear that this can be explained by adequate classroom management, the segment should be scored as a 4.
  - There can be a disruption (e.g. loud voices in the hallway, a student entering the classroom late), and the segment can still be a rated a 4 if instructional time is not lost.
- Evidence: teacher and students.

Table A B.4. Time on task

Indicator	1	2	3	4
Time on Task. Most of the segment for the "ideal student" is focused on mathematical learning. There is little loss of lesson time to activities or situations that are not directly focused on mathematical learning (e.g. greeting one another, behaviour issues, classroom routines, transitions, off topic discussions). Mathematical learning can be full range of activities in which the ideal student should be engaged – listening to a lecture, doing group work, working alone on a problem, etc.	50% of the segment or more (i.e. usually 4 minutes or more) is lost to activities, tasks, or dialogue nor focused on mathematical learning.	Between 25% and 49% of the segment (i.e. usually 2 minutes to 3 minutes 59 seconds) is lost to activities, tasks, or dialogue nor focused on mathematical learning.	Between 24% and 7% of the segment (i.e. usually 30 seconds to 1 minute 59 seconds) is lost to activities, tasks, or dialogue nor focused on mathematical learning.	6% of the segment or less (i.e. usually 30 seconds or less) is lost to activities, tasks, or dialogue nor focused on mathematical learning.

Definition of time on task

84. Time on task is the amount of time the group of students spends on mathematics or mathematical activities.

- 1: 50% of the segment or more (4 minutes or more) is lost to tasks and activities not focused on mathematical learning.
- 2: Between 25% and 49% of the segment (2 min to 3 min 59 secs) is lost to tasks and activities not focused on mathematical learning.
- 3: Between 24% and 7% of the segment (30 seconds to 1 min 59 secs) is lost to tasks and activities not focused on mathematical learning.
- 4: Less than 6% of the segment (30 seconds or less) is lost to tasks and activities not focused on mathematical learning.

Notes and rating guidance

- Any time that the classroom of students does not spend on mathematics should be accounted for in this code.
- We are not trying to capture differences within a classroom of students in this code. Teachers often ask students to do something and not all students do it at the same pace. For example, a teacher might ask a class of 25 students to take out their books and turn to page 65. Most students might do this, but the rater could notice 2 students who continue to chat and get out their books after the other students. The rater should pay attention to what the whole class is doing, not those two students.
- In cases where many students (half or more of the students in the classroom) do not do what the teacher has asked or are no longer engaged in a mathematical activity, the rater should count loss of time on task.

- In this code, we are most interested in the time spent on any mathematical activity – the importance or quality of that activity should NOT be judged in this code. If the class is carrying out a mathematical activity, this counts as time on task.
- E.g. If the teacher was talking about an “IB test” (a test that is not part of the lesson but one that the students will all take), and the teacher was explaining that when taking such a test, the student should think about how to communicate clearly about the mathematics with the test’s assessor – using descriptive words and writing all equations carefully, this would count as time on task. This counts because the teacher was talking about general test taking skills in math. If the teacher was talking about getting a good night sleep the night before, we would count that as wasted time.
- Behaviours such as a pledge, prayer, and greetings/bows at the beginning of the class will be counted as loss of instructional time.
- Evidence: teacher and students.

Table A B.5. Activity structure and frequency

Indicator	Arrangement of students in grouping structures	
Activity structure and frequency. Check which structures were used and how frequently. Briefly = less 50% or less than 4 min More than briefly = 50-99% or 4-7:59 min Used the entire segment = the whole segment or 8 min	1. Whole group	Not used (1), Briefly used (2), more than briefly used (3), used the entire segment (4)
	2. Small group (3+)	Not used (1), Briefly used (2), more than briefly used (3), used the entire segment (4)
	3. Pairs	Not used (1), Briefly used (2), more than briefly used (3), used the entire segment (4)
	4. Individual	Not used (1), Briefly used (2), more than briefly used (3), used the entire segment (4)

Definition of activity structure and frequency

- Definition of activity structure:
  - Individual work: teacher is silent or helping students individually; students carry out mathematical work individually, often at their seats.
  - Pairs: teacher is silent or helping students in groups of two; students carry out mathematical work in groups of two.
  - Small group: teacher is silent or helping students in groups of three or more, students carry out mathematical work in groups of three or more.
  - Whole group: teacher addresses students as a whole class; mathematical work is carried out as a whole group.
- Definitions of length of use:
  - Not used: activity structure is not used in the segment.
  - Briefly: activity structure is used for less 50% or less than 4 min.
  - More than briefly: activity structure is used for 50-99% or 4-7:59 min.
  - Used the entire segment: the whole segment or 8 min.

Notes and rating guidance

- When rating, begin with the structure the teacher has asked students to use (e.g. individual or pair work), unless you have strong evidence to the contrary. "Strong evidence to the contrary" is when more than half of the students are using a different structure than indicated by the teacher.
- If there is no explicit statement about what grouping the students should use, the rater should go back to when the activity began and try to determine what grouping was implied by the teacher.
- If it is difficult to tell what structure was implied or there are different structures are being used by students (e.g. 2/3 of the class is working individually but 1/3 is working in pairs) the rater should use the following rules:
  - Determine what structure half or more of the class is using and use that as the code for all of the class.
  - If the class uses a mixture of structures and none of them reach the level of "half or more of the class", code the largest sized group activity structure being used.
- The rater will need to pay attention to timestamps on the video to rate this correctly.
- When students are working at their seats individually and the teacher addresses all of them, that counts as whole group. There will be two things coded when that happens: individual work and whole group. The same principle applies if the teacher addresses the whole group while students are working in pairs or in small groups. Both should be coded depending on the amount of time used for each.
- Individual work is not always silent. Students often talk to their neighbour during individual work; this does not mean they are working in pairs or groups. Pay careful attention to the structure stated or implied by the teacher. If the norm appears to be that students are allowed to talk to one another during individual work, this should still be counted as individual work.
- Mental shortcut: focus on half of the segment.
- For "down time" when there is no instruction going on and students are not working on mathematics, code the activity structure as "individual", unless the teacher is addressing the students as a group, for which we will code "whole group".
- Specific examples:
  - When the teacher is writing or has something written at the front of the room and students are taking notes, this is considered whole group.
  - If the teacher says "go ahead and get started on your worksheet" and the students begin to work quietly individually, this should be coded as individual work.
- Evidence: teacher and students.

Table A B.6. Time of lesson

Time of lesson	Start time of lesson	End time of lesson
This is the length of the lesson from the beginning to end of the digital recording.	0:00:00	

Social-emotional support

Table A B.7. Respect

Component	1	2	3	4
Respect. Teacher and students demonstrate respect for one another by using any of the following types of behaviours: respectful language, listening to one another, using appropriate names, using a respectful tone of voice, and using traditional markers of manner.	Teacher and students rarely demonstrate respect for one another.	Teacher and students sometimes and/or inconsistently demonstrate respect for one another.	Teacher and students frequently demonstrate respect for one another, although there may be inconsistencies.	Teacher and students frequently and consistently demonstrate respect for one another.
There are no disrespectful interactions between the teacher and students, or between students (i.e. threats, mean or degrading comments, physical aggression such as pushing someone or slamming down materials, comments after which student or teacher demonstrates shame).	There are a few brief and/or minor negative interactions or one sustained and/or substantial negative interaction between any student and the teacher, or between students.	There are 1-2 brief and/or minor negative interactions between any student and the teacher, or between students.	There are no negative interactions between any student and the teacher, or between students.	There are no negative interactions between any student and the teacher, or between students.

Definition of respect

85. Because respectful behaviours vary across countries, TALIS Video participants have had to specify a relatively small set of respect and disrespect behaviours that will be "counted" across countries in order to apply this code in a standardised way.

86. Raters should look for and rate ONLY the following five specific behaviours:

1. respectful language
2. listening to one another (e.g. nodding, making eye contact, waiting for another speaker to finish speaking)
3. using appropriate names
4. using a respectful tone of voice
5. using traditional markers of manners (i.e. please and thank you).

Notes and rating guidance

- Interpreting certain behaviours could vary across countries. For example, in some countries students are required to raise their hands and get teachers' signals before they are allowed to speak publicly. In those contexts, such behaviours may not necessarily be interpreted as respect, but rather an established routine or norm.

- Disrespectful interactions include: threats, mean or degrading comments, physical aggression such as pushing someone or slamming down materials, comments after which the student or teacher demonstrates shame.
- If there is no disrespectful language or use of a disrespectful tone of voice, the rater should presume the language and tone are respectful. The rater should pay attention to the reactions of students and the teacher in the room to determine what is respectful and disrespectful.
- Rating guidance:
    - An example of disrespectful interactions between the teacher and students:

      T: Do we shut up now? Or do I start yelling? Good. If you talk again I want it to be clear I'll start penalising you. Did you all hear me? If you talk again, I'll start penalising you. Is that clear? Good. Sit properly and shut up; and copy from the blackboard right now.

        - Consistency: raters should evaluate the consistency in each of the respectful behaviours present. To be scored at a 4, each of the respectful behaviours being observed should happen consistently throughout the segment; whereas for the score of 3, maybe one or two respectful behaviours are not as consistent as the others.
        - When the teacher has to prompt students to listen to one another, this would be considered a lack of consistency in demonstrating "listening to one another".
        - Raters should pay careful attention to the benchmarks in training to distinguish between score points.
        - Student to student respect behaviours should also be counted.
- Evidence: teacher and students.

Table A B.8. Encouragement and warmth

Component	1	2	3	4
Encouragement and warmth. The teacher and/or students provide encouragement to students throughout their work. (i.e. the teacher may reassure students when errors are made, make positive comments, compliment students' work.)	The teacher and/or students do not provide encouragement to students throughout their work.	The teacher and/or students occasionally provide encouragement to students throughout their work.	The teacher and/or students sometimes provide encouragement to students throughout their work.	The teacher and/or students frequently provide encouragement to students throughout their work. .
There are moments of shared warmth (i.e. smiling, laughter, joking and playfulness).	There is no evidence of shared warmth.	There are occasional moments of shared warmth.	There are some moments of shared warmth.	There are frequent moments of shared warmth.

Definition of encouragement and warmth

- Encouragement: positive verbal and/or nonverbal cues that may inspire or motivate students to begin or keep trying to accomplish a task. Examples: reassuring students when errors are made, complimenting students' work, making positive comments Encouragement provided by students should be counted.
- Warmth: shared warmth between the teacher and students, and among students. Examples: smiling, laughter, joking, playfulness.

Notes and rating guidance

- Raters should look for only the behaviours listed in the definitions and rate based on those behaviours.
- If there is a flat affect between the teacher and students, it should be accounted for in encouragement and warmth.
- Examples of encouragement:
  - Compliment students' work

    T: Negative 1. Over…

    S: 2 times -1/2.

    T: Very good, Sofía. Times -1/2. Okay.
  - Whole class encouragement and shared warmth
    - After a student presented his work,

      T: Congratulations! Give him a hand!
- Evidence: teacher and students.

Table A B.9. Risk-taking

Component	1	2	3	4
Risk-taking. Students seek guidance.	Students do not seek guidance.	Students rarely seek guidance.	Students sometimes seek guidance.	Students frequently seek guidance.
	And/or	And/or	And/or	And/or
Students voluntarily take risks by publicly sharing their private work.	Students do not voluntarily share their private work publicly.	Students rarely voluntarily share their private work publicly.	Students sometimes voluntarily share their private work publicly.	Students frequently voluntarily share their private work publicly

Definition of risk-taking

- Risk-taking: The extent to which students are willing to share their thinking with the class voluntarily or ask questions of the teacher or their peers publicly. Taking risks is an internal process for the student.
- Guidance: When a student communicates a need for information, advice, or help from the teacher. This could include asking questions about the student’s work, making a facial expression indicating confusion. It could happen in front of the whole class, in a small group, in pairs, or between a teacher and a single student.
- Voluntarily sharing: When students share their private thinking without the teacher specifically calling on a particular student to respond, it is voluntary sharing.
- Sharing private work publicly: Whenever students make their internal thinking or problem-solving process available for their peers to read or hear. For example, a student may write the solution to a problem and the step(s) he took to solve it on the whiteboard for the whole class to see.
- Distinction between “public” and “private”: Public refers to sharing student thinking in front of the whole class; private work refers to what is going on in students’ minds or on their papers in front of them.

Notes and rating guidance

- In this component, we only measure students’ observable risk-taking behaviours--seeking guidance and publicly sharing their private work.
- Risk refers to emotional risk; it could be a risk to one’s reputation or self-esteem. It might also be subjecting oneself to evaluation by the teacher or peers.
- Rating guidance:
    - When the rater is deciding between two rating levels, refer back to the idea of emotional safety that underlies the domain.
    - Do not count choral responses in this code because our goal is to capture the risk associated with speaking publicly when one’s voice can be recognised as one’s one. Be careful when you discount choral evidence, however. Sometimes there is a lot of choral response but a few times (within the choral response) when a single voice can be heard alone. Those single responses should be counted as risk-taking.
- Evidence: teacher and students.

Table A B.10. Persistence

Indicator	1	2	3	4
Persistence. Student(s) persist through errors or mathematical struggles with the teacher's support.	Student(s) did not make errors or engage in mathematical struggle.	Student(s) are aware they have made an error or are engaged in a mathematical struggle.	Student(s) are aware they have made an error or are engaged in a mathematical struggle.	Student(s) are aware they have made an error or are engaged in a mathematical struggle.
	Or	And	And	And
	Student(s) are not aware they have made an error or are engaged in a mathematical struggle.	Mathematical errors or struggles are either ignored or briefly and/or superficially addressed.	Student(s) persist through mathematical errors or struggles in a moderate length and/or depth with the teacher's support.	Student(s) persist through mathematical errors or struggles in significant length and/or depth with the teacher's support.

Definition of persistence

87. Persistence is a student <u>or</u> students show effort to address an error or struggle over time.

Notes and rating guidance

- This code belongs to the social-emotional domain. The rater's focus should be on the students' emotional experience of error and/or struggle.
- Engagement in cognitively challenging mathematics is NOT what is measured in this code. The rater should identify errors of which the student is aware or the presence of (emotional) struggle.
- The error or struggle that is coded may be for a single student or a group of students or the whole class.
- We presume that if a student keeps struggling through difficulties (even if the teacher does not intervene) there is the presence of persistence.
- If problems or difficulties are resolved by the individual student who made the error OR at the class level, they are considered to be addressed.
- The rater should first ask: Is there an error or struggle? And is this error or struggle about the mathematics of the lesson? If yes, to what degree did the students continue to show efforts to address the error or struggle?
- Examples of struggle include but are not limited to:
  - S: "I don't understand." "I am confused." "I get lost here." "This is too hard!" "I can't do this."
  - There are repeated efforts to understand or solve a problem that have demonstrated emotions (e.g. frustration, anger, confusion) associated with those efforts. This could look like a student or group of students showing they do not understand, seeking teacher guidance, following the guidance the teacher provides, not being successful in reaching the correct answer AND showing confusion and/or frustration with their inability to reach the correct answer.
- Examples of errors of which the student is aware include but are not limited to:
  - The teacher states or notes physically (e.g. pointing) that the student's answer, procedure, or thinking is incorrect or problematic.
  - The student states or writes that s/he made a mistake, got something wrong, or carried out the mathematics incorrectly.

- The following are generally NOT evidence of error or struggle:
  - Students asking if they are on the right track.
  - Students clarifying or asking the teacher to check the correctness of procedures or answers they have developed.
  - Students' general questions or uncertainty.
  - Working on an activity or problem for a long time.
  - A teacher being highly prescriptive about certain aspects of the activities (for example, how students are formatting or decorating a poster, the specific mathematical language the teacher prefers the students use, the teacher reminding students about how to name a mathematical idea, procedure or process, etc.).
- A rating of 1 will be given if there is no error or struggle. This does not indicate a low level of persistence generally for the teacher and students, but rather, there is no evidence of persistence shown in the specific segment being rated.
- There is a difference between a student simply reporting an error and a student exhibiting evidence that they understand they made an error or saying they are struggling.
- If the student has already resolved the error and is reporting that they made that error but have resolved it, that situation is not eligible for persistence because the error has happened in the past. The examples below should not be counted as evidence of persistence.
  - (Student raises hand)

    S: Teacher, I just realised I made the same mistake Gabriella and Ben made. But I fixed it.

    T: Ah good! You have to be very careful when you substitute. It is easy to make a sign mistake and the whole problem will be wrong.
  - (Teacher circulating among the students)

    T: How are you doing Naoko?

    S: I think I understand. At first I did not know how to represent the relationship between the length of the sides of the rectangle and the maximum area, but Kenji explained it to our good and we understand.

    T: Good. Can you show me?

    (Student explains)
- Evidence: teacher and students.

Table A B.11. Requests for public sharing

Indicator	1	2	3
Requests for public sharing. The teacher requests students share their private mathematical thinking publicly. Note: Small group work is NOT considered public sharing.	The teacher does not request students share private mathematical thinking publicly. If there is a request, it is rare and the shared work has limited detail.	The teacher requests students share private mathematical thinking publicly. Shared work has limited detail.	The teacher requests students share private mathematical thinking publicly. Some shared work has more than limited detail.

Definition of requests for public sharing

88. The focus of this code is on the degree to which teacher's request students participate in the mathematical conversation of the classroom. We count any type of public sharing in this code.

- Definition of terms:
  - Limited detail: This describes student contributions that briefly name an answer but do not reveal students' thinking processes or rationales.
    - Examples: Naming a numeric answer; naming a mathematical term; providing a definition of a term or process.
  - More than limited detail: This describes student contributions that reveal students' thinking processes or rationales.
    - Examples: Explaining or describing procedures taken or reasoning behind procedures taken.
  - Distinction between "public" and "private": Public refers to sharing student thinking in front of the whole class; private work refers to what's going on in students' minds or on their papers in front of them.

Notes and rating guidance

- It does not matter if the teacher made the request to the whole class or toward specific students. The focus in this code is on the request resulting in public access to students' thinking.
- Choral responses can count for this code. They can count because this code is trying to determine the degree to which students contribute to the discourse and what types of thinking are contributed. In other codes we are stricter in the degree to which the students' talk must represent a social-emotional risk.
- The word "some" in the 3 descriptor "Some shared work has more than limited detail" requires two or more instances of "more than limited detail".
- The word "rare" in the 1 descriptor allows the teacher to request student thinking once without moving the rating up to a 2.
- Evidence: teacher and students.

Discourse

Table A B.12. Nature of discourse

Component	1	2	3	4
Nature of discourse. Students have opportunities to participate in the classroom discourse.	Discourse is teacher-directed.	Discourse is frequently teacher-directed.	Discourse is sometimes teacher-directed.	Discourse is rarely teacher-directed.
Students' discourse is characterised by detailed contributions.	Students' discourse does not include any detailed contributions.	Students' discourse is rarely characterised by detailed contributions.	Students' discourse is sometimes characterised by detailed contributions.	Students' discourse is frequently characterised by detailed contributions.

Definition of nature of discourse

- Discourse: Discourse is any communication in the classroom by the teacher and/or students. Written discourse (e.g. words and symbols on the board or on worksheets at students' desks) is considered discourse.
- Detailed contributions: Contributions that have sufficient detail about the mathematics being worked on, not just short answers that give the answer or define a term, for example. This code accounts for the level of specificity and information provided in the students' contributions. Detailed contributions tend to be longer than less detailed contributions.
- Teacher-directed discourse: Communication in which the teacher has control over the pattern of questions and answers. This includes teacher lecture as well as student-centred conversations in which the teacher initiates a question, a student responds, the teacher evaluates the correctness or completeness of the response and then the teacher begins the cycle again with a new question or statement. In teacher-directed discourse, students may pose questions, but they do not substantially shape the direction or nature of the mathematical discourse.

Notes and rating guidance

- Only discourse around mathematics will be counted as evidence for this component.
- Choral responses should be taken into account on this component.
- Rating guidance:
  - When students are working with small groups or in pairs, students are often directing the discourse, and the segment will rate higher on the first descriptor.
- Examples of detailed contributions:
  - T: How did you do number 7? Will?

    S: I started by moving the 24 to the left side of the equation so that everyone was equal to zero. Then I factorised, then solved.
  - When students are asked to solve the following problems on the worksheet in front of them:

    X2+4x-2 = 0

    2x2+2x = 0

3x2 – 69 = 0

S: The conclusion is "x (10 - x) = 16."

T: OK.

S: The length is x, so subtracting 2x from the perimeter of 20 m gives the value of 2 sides of width which we divided by 2. Then, the value of 1 side of width can be expressed as 10 minus x. Expanding the expression on and on leads to "x2 - 10x + 16 = 0."

- When students are divided into groups and asked to work together to represent and solve the following word problem:
  - The perimeter of a rectangle is 20. What is the area of the rectangle?

- Evidence: teacher and students.

Table A B.13. Questioning

Component	1	2	3	4
Questioning. Questions request students engage in a range of types of cognitive reasoning.	Questions generally request students recall, report an answer, provide yes/no answers, and/or define terms. Students' discourse does not include any detailed contributions.	Questions generally request students recall, report an answer, provide yes/no answers, and/or define terms although there are some questions that request student summarise, explain, classify, or apply rules, processes, or formulas.	Despite a few questions that request students recall, report, and /or define, most questions request that students summarise, explain, classify, or apply rules, processes, or formulas. There may be a small number of questions that request students analyse, synthesise, justify, or conjecture.	Questions request a mixture of recall, reporting, defining, summarising, explaining, classifying, applying rules, processes, or formulas, analysing, synthesising, justifying, and/or conjecturing, but the emphasis is on questions that request students analyse, synthesise, justify, or conjecture.

Definition of questioning

89. Questioning focuses on the nature of questions teachers ask in the classroom.

Notes and rating guidance

- This code solely considers the nature of the questions that teachers ask – in written and oral forms. Students' answers should be taken into account to determine the nature of the question, but the nuance and meaning of students' answers should be considered in the Assessment of and Responses to Student Understanding Domain.
- The rater must look at student worksheets and/or problems carefully so that those questions are also accounted for in the rating.
- Rhetorical questions (i.e. questions the teacher poses and either does not answer or answers him or herself) should not be counted toward the questioning rating.
- The rater should ask: What kinds of questions characterise the segment?
- Examples of questions that request students recall, report an answer, provide yes/no answers, and/or define terms:
  - (After solving an incomplete quadratic equation) T: What did you get Patrick?
  - T: What is the equivalence principal?

  - T: What is a? B? C?
  - T: Did you understand that explanation?
  - T: Do you remember what we did yesterday?
- Examples of questions that request students summarise, explain, classify, or apply rules, processes, or formulas:
  - T: Can you tell me, how did you get this answer?
  - T: Let's see if substituting 4 and 8 each into x in equation 2 would work. Why do we substitute 4 and 8?
  - T: How many conditions do the roots of quadratic equations with one unknown have? What are they?
- Examples of questions that request students analyse, synthesise, justify, or conjecture
  - The perimeter of a rectangle is 20. What is the area of the rectangle?
  - What is the pattern you notice across the three problems we just solved? Look carefully.
  - Jon, can you explain why you disagree? Why do you think completing the square is a more efficient approach than just using the quadratic equation for number 4 on the board?
- Rating guidance:
  - At a 4, the questions should generally be questions focused on analysis, synthesis, justification, or conjecturing.
  - When a student takes a word problem and creates a diagram or graph from it, this is analysis.
- Evidence: teacher and students.

Table A B.14. Explanations

Component	1	2	3	4
Explanations. Teacher and students provide written and/or verbal explanations. Explanations are descriptions of why ideas or procedures are the way they are.	There are no explanations of why ideas or procedures are the way they are by either teacher or students.	Explanations generally focus on brief and/or superficial features of the mathematics.	Explanations focus on a mixture of brief/ superficial and lengthy/deeper features of the mathematics.	Explanations focus on lengthy/deeper features of the mathematics

Definition of explanations

90. Explanation is a description of why ideas or processes are the way they are. It is a statement that clarifies, rationalises, and/or justifies.

Notes and rating guidance

- It is very important for the rater to determine if what is being said focuses on why ideas or processes are the way they are. There are many times teachers and students speak and there is little or no emphasis on why.
- Explanations can come from the teacher or the students and can be stated in whole group, small group, or individual activities.
  - An example of a brief and superficial explanation:

    T: "Finally the last one is called a constant. Why is it called a constant?"

    S: "It doesn't change."

    T: "Perfect. It doesn't change. No matter what, it's always constant, always the same."
- Evidence: teacher and students.

Table A B.15. Discussion opportunities

Indicator	Note: Discussions are extended conversations between and among the teacher and many students where students do much of the talking. Though the teacher guides the discussion towards a learning goal, discussions are predominantly based on student ideas and characterised by student-to-student interaction.
Discussion opportunities. Whether a segment of instruction engages students in discussions that are clear and focused on the learning objective.	1. Not present 2. Present

Definition of discussion opportunities

- Discussion: extended conversations between and among the teacher and many students toward a learning goal. Although the teacher guides the discussion towards a learning goal, discussions are predominantly based on students' ideas and characterised by student-to-student interaction.
- Learning objective: The goal for student learning around which the lesson is based.

Notes and rating guidance

- There must be both an extended conversation with students speaking to one another (or one another's ideas) and that conversation must be directed toward a learning goal by the teacher in order to be counted in this code.
- Only discussion around mathematics is counted as evidence for this indicator.
- If the discussion occurs over multiple segments but the rater does not realise the discussion qualifies as a discussion until a later segment, the rater should determine the segment in which the discussion started and then change the rating on those segments (to a 2). The need to change codes should be rare. Discussions are defined in such a way that it generally should be obvious to the rater that the students are having a discussion. Pay careful attention to the definition of discussion.
- Example: Teacher hands each student a card with a quadratic equation written on it (e.g. one student has 2x2-x+1=3; another student has 3x2=5). Students are then asked to come up to the front of the room and tape their card into one of four sections of the white board – each student must sort his or her example of a

quadratic equation based on the general form they match: (1) ax2+bx+c=0 (2) ax2=0 (3) ax2+bx =0 or (4) ax2+c=0. After sorting the examples, the teacher and students engage in the following whole group interaction.

T: What structure do you believe to be most recognisable?

S: The general one.

T: The general one? The ones that have?

S: The ones that have the same value in the exponent.

T: "Value in the exponent"? If we look closely, all of them will have the same exponent in at least one of the terms.

SN: The orange one over there is wrong.

T: You say the general form is easiest. Because you understand the general formula best? Or because it is, in fact, the easiest one to identify.

SN: I just chose it randomly.

SN: It's the one we started with.

T: What are your criteria in classifying the easiest formula? There's no wrong answer, but what criteria do you have in mind before saying which is the easiest one?

SN: It should help me identify "ax2". Then "bx" that is, "x" without the exponent. And "c". It should have a value for "c", regardless of what that is.

T: Now, who thinks ax2=0 is the easiest one? (some Ss raise their hands) David, why?

SN: Well, because, as you can see we can see the variable is raised to the same exponent so we have a group of signs.

T: A group of signs…

S: I mean, terms.

T: Terms…So maybe you're thinking of these over here (pointing to examples students have sorted under ax2=0). So Maggie was looking for one that had a term with a squared variable, one with no exponent, and an individual number. Hence justifying his reason to look for that particular form instead of the others. It was the easiest one for him…Is that clear?...Now, do we follow blindly? Has he convinced us? Is that really the easiest form to identify? Or isn't it? Luis?

SN: ax2=0 is obviously the easiest one because it only has the ax2 term and nothing else.

T: That's it. It's fine. In his case, I would have chosen the same one. Me. I speak for me. But Erika's answer wasn't wrong. Why? Because she's taking into account every possible term we can be given. The complete general structure, right? …Now we will learn that it's not only about classifying the structure it has. This will also lead us to a way of solving each equation…

T: Moving along, following Luis's criteria, what would the [next easiest] structure be?

SN: ax2+bx=0

SN: either one

T: ax2+bx=0 or ax2+c=0?

SN: I would choose ax2+bx=0.

T: Why?

S: Because well in my case, I would guide myself with the order of the general equation.

T: The order of the general equation.

S: So because the general form is ax2+bx+c, I would follow that order. Therefore, the next structure would be ax2+bx.

T: But, what if it's not following that particular order? I mean that's ok, it's a valid argument. The thing is, you adhere to that strategy, but what if the equation is not in the same order? What would you do? Felipe?

SN: He already said that he's following the order of the general structure.

T: Mariam?

SN: I would go with ax2+c=0.

T: Why that one?

S: Because it has only one literal coefficient. The same literal coefficient of ax2=0.

T: it has a literal coefficient. Ok yes. Also it has only one variable. Speaking of literal coefficients, "C" is also a literal coefficient. What is the literal coefficient in c?

SN: x to the…

T: x to the…x to the zero power, right? There it is. Meaning if we take this, we have x2, x1, and x0, because x0 equals 1, right? Is that clear? Now as he points out, both (ax2+c=0 and ax2+bx=0) are somewhat similar right? What's different about them?

SN: there's an independent term.

T: Exactly, there's an independent term. Regardless, let's see if we classified them correctly.

<Teacher continues to lead the students in evaluating whether or not each example is sorted correctly>

- The exchange above exemplifies several features of discussion. First, the students' ideas guide the direction of the discourse. For example, the teacher references "Luis's idea" as one way to sort quadratic equations. Second, the teacher is shaping the discussion with the learning goals in mind. His questions focus on how to sort quadratic equations, but he mentions the idea that they will build upon this skill when they learn how to solve each type of equation. Third, the teacher asks students to respond directly to one another. For example, he asks, "Has he convinced us? Is that really the easiest form to identify?" Fourth, the entire conversation lasts several minutes so it is considered an extended conversation.

- Evidence: teacher and students.

Quality of subject matter

Table A B.16. Explicit connections

Component	1	2	3	4
Explicit connections. Teacher or students make explicit instructional connections between any two aspects of the subject matter.	There are no instructional connections between ideas, procedures, perspectives, representations or equations.	There is one instructional connection between ideas, procedures, perspectives, representations or equations.	There are at least two instructional connection between ideas, procedures, perspectives, representations or equations.	There are at least two instructional connection between ideas, procedures, perspectives, representations or equations.
	Or	And	And	And
Aspects include subject matter ideas, procedures, perspectives, representations or equations.	Connection(s) that are present are implicit.	Connection(s) are generally explicit, but vague.	Connection(s) are generally explicit, clear and brief	Connection(s) are explicit and clear, and at least one is elaborated.

Definition of explicit connections

- Instructional connection: A relationship or association that is called out by a teacher or student.
- Explicit: Stated clearly, verbally or in written form.
- Implicit: Implied, not specifically or clearly stated.

Notes and rating guidance

- Explicit instructional connections should be counted within the topic of quadratic equations. Instructional connections are between and among various aspects of the mathematics. Aspects are subject matter ideas, equations, representations, perspectives or procedures.
- Explicit instructional connections between quadratic equations and mathematical topics outside of quadratic equations or the real world also can be counted if they concern ideas, equations, representations, perspectives or procedures in those topics and real-world settings.
- Instructional connections may be elaborated or brief. Elaboration often, although not always, coincides with the connection spanning more time in the lesson.
- As you consider the instructional connection, consider the degree to which the nature of the connection is specified in detail and made instructionally visible. It is not enough to state that there is a connection. What the connection is must be specified in some detail.
- More detail for both aspects being connected (e.g. the equation and the idea) is likely to result in a connection that is counted as elaborated.
- Details can be provided for any aspect of the mathematics – regardless of the importance of the mathematics, or the surface or deep nature of the mathematics. This results in the possibility that instructional connections may be elaborated about surface features of the mathematics.
- This code does not measure mathematical connections that are present regardless of what the teachers or students say or do. Many of these types of connections are

definitional in nature or are opportunities for the teacher to name the mathematics the students are working on.

  - For example, when a teacher asks a student to graph a quadratic equation there is a connection between the graph and the equation. This would NOT count unless the teacher makes the connection explicit and draws attention to it.

- Raters need to identify and **count** the instructional connections and rate based on the rubric.
- When students are making substitutions from one equation into another (e.g. from a problem to the quadratic equation), this NOT counted as a connection.
- See the handout for more examples of connections.
- Perspectives on quadratic equations include:
  - conic sections
    - A conic sections perspective (as contrasted with a perspective that uses Cartesian coordinates, or the algebra you're used to seeing). [1] From a conic sections perspective, the quadratic function is a type of parabola (note that while teachers and students use the terms interchangeably, strictly speaking quadratic functions are a subset of parabolas) defined as the curve at the intersection of a cone and a plane. Parabolas are one of four possible types of intersections, and they occur when the plane is parallel to one of the "sides" of the cone (see wiki link for pic). Parabolas are defined by the relationship between a focus point and a line called the directrix, and you can think of the set of points on the parabola as being the set of points equidistant from the focus and the directrix. The equations representing this relationship are convertible to Cartesian coordinates (x, y) but are not equivalent to them; they represent a different way of parameterising the curve. The study of conic sections is the historical origin of parabolas, but is rarely taught in many countries' mathematics curricula, with the following two exceptions that might be observed in lessons. First, sometimes teachers will provide a physical demonstration with a cone of how the slicing with a plane produces different curves, using a foam cone or sometimes a beam of light in a cone shape.[2] Second, teachers who use a programme like geometer's sketchpad may use the locus of points functionality to have students discover or define a curve based on a particular locus relationship.[3]
      - **What would this look like?** It is likely you will almost never see it but if you do, it is likely to stick out. Teachers or students would say things like "a cone sliced by a plane" or talk about the focus and directrix.
      - **Connections:** Explicit algebraic connections can be made between the two methods of representing the quadratic function.
  - quadratics as polynomials

[1] See, https://en.wikipedia.org/wiki/Conic_section for an overview and picture

[2] See, https://divisbyzero.com/2009/03/11/flashlights-and-conic-sections/

[3] See, http://jwilson.coe.uga.edu/EMT668/EMAT6680.F99/Challen/iu/day1.html

- Thinking of quadratics as sort of their own free-standing topic versus thinking of quadratics as polynomials of degree two, a subset of the larger class of polynomial functions.
  - **What would this look like?** This first perspective might be implicit in a focus on methods that are particular to quadratics, like completing the square or the use of the quadratic formula, or if graphing, a focus on use of the vertex formula or finding roots and using symmetry. The second perspective might be visible in the use of methods that work more generally, such as the use of the factor theorem with synthetic division to factor a polynomial with rational roots or the use of an approximation method like Newton's method[4] to find any roots, or if working with graphs, using function transformation as a framework. Or in making specific comparisons to other families of functions, like pointing out that first differences are constant for linear functions, second differences for quadratics, or calling attention to the degree as giving you the number of (complex) roots in all cases, or the number of "bends" in the curve, etc.
  - **Connections:** The most obvious connection would be to other families of functions.

o quadratics as representations of real world phenomena

- Quadratics as representation of real world phenomena vs as pure mathematics. A number of physical phenomena are quadratic in nature. The obvious example is gravity (which gives you bottle rockets and throwing a ball in the air and that whole family of experiments). Note that this is more than just "stuff that can be adequately modelled by quadratic functions" as described below. Teachers could take that approach, but gravity is more than just adequately modelled by quadratic function, its equation (absent friction) is a second-degree polynomial. It is an exact fit. This perspective includes the kinds of word problems in textbooks that give a distance function and the student has to find the time when some object hits the ground (and other similar problems).
  - **What does this look like?** Teachers and students talk about the real-world phenomena; for example, height of a ball thrown in the air, or assign that type of problem.
  - **Connections:** This has natural ways of connecting to representations of the real-world phenomena.

o quadratics as best fit model

- Quadratics as one of a number of models to approximate a set of data and/or real-world phenomena. This would include looking at data and deciding if it is quadratic in nature (either exact, or approximate, as in a good enough fit). This might involve looking at graphs, or examining second differences, or learning how to run regressions on various technology or software. It also likely involves explicit comparison to other function types, like linear, cubic, or exponential models, that might also be considered in terms of fitting the data.

[4] See, https://en.wikipedia.org/wiki/Newton%27s_method

        - **What does this look like?** Working from data
        - **Connections:** There are a lot of connections available here to real world phenomena perspective, to ways of representing the data (table, graph, formula), etc.

- Examples of connections:
    - Implicit connection: (procedure to procedure)

      T: (to whole class) Do you remember that the last time we were together we worked on solving quadratic equations by inspection?

      S: Yes.

      T: Today we'll be solving them by completing the square. Open your books to page 34.

      More explicit connection: (representation – an apple flying through the air to a mathematical idea – parabola)

      T: Lisa has an apple. And she throws. And then there is a certain kind of curve. Can you imagine something? The apple flies and how does the mathematical curve look like?

      S: Parabola.

    - Explicit (and brief) connection: (representation – FDX=3x+2, x to the first degree and representation – straight line)

      T: I have the following function, FDX equals 3X plus two. As the X is degree one, it means this – that this is a linear equation, as the name says, it is a straight line, it is a line.

- Connections must be explicitly stated to rate above a 1.
- Evidence: teacher and students.

Table A B.17. Explicit patterns and generalisations

Component	1	2	3	4
Explicit patterns or generalisations. The teacher or students look for patterns in their work together. They also generalise from the specific work the students are working onto a foundational concept and/or definitions underlying the specific work.	Neither the teacher not students look for patterns in the mathematical work.	Teacher looks for patterns in the mathematical work. Identified patterns focus on surface features of the mathematics.	Students look for patterns in the mathematical work. Identified patterns focus on surface features of the mathematics.	Teacher or students look for patterns in the mathematical work. Identified patterns focus on surface features of the mathematics.
	Or	Or	Or	Or
	They do not generalise from the work.	Explicit generalisation(s) are developed from the mathematics under consideration and focus on nomenclature or algorithm processes. They are muddled, correct or incorrect, and superficial.	Explicit generalisation(s) are developed from the mathematics under consideration and focus on nomenclature or algorithm processes. They are clear, correct, and elaborated. If they generalise to foundational concepts, ideas, and/or definitions, the generalisations are somewhat muddled.	Explicit generalisation(s) are developed from the mathematics under consideration and focus on nomenclature or algorithm processes. They are clear and correct.

Definition of explicit patterns and generalisations

- Pattern: An ordered set of mathematical objects (e.g. numbers, equations, graphs, problems), a recurring sequence.
- Generalisation: "Generalisation involves deliberately extending the range of reasoning or communication beyond the case or cases considered, explicitly identifying and exposing commonality across cases, or lifting the reasoning or communication to a level where the focus is no longer on the cases or situations themselves, but rather on the patterns, procedures, structures, and the relations across and among them (which, in turn, become new, higher level objects of reasoning or communication)." (Kaput, 1999[55])

Notes and rating guidance

- There must be at least two examples referred to or investigated from which generalisation or pattern is developed.
- Students must be explicitly asked to look for the pattern.
- The rater should attend to five aspects of patterns and generalisations: 1) who notices the pattern, 2) the presence of patterns; 3) the pattern's quality; 4) the presence of explicit generalisations; and 5) the generalisation's quality.
- Examples of patterns and generalisations:
  - T: We saw two patterns of explanation. Here. All groups tried to get the value of width. I think all of you share this approach, right? … Apply the distributive law, you all did this, and transpose the term, you all did this too.
  - T: (After completing three problems) Now this is for any function, linear function and quadratic function, always the free number of X indicates where the cut of the function is – where the Y-axis cuts.

- Rating guidance:
  - Statements that are made without prior development of at least two examples do not count.
  - Sometimes teachers ask students to think about numbers that "when multiplied by zero will give you zero". Unless there is clear evidence students are actually going through the process of multiplying least 2 different numbers by zero before making a conclusion, this should not count as a pattern.
  - Examples of general statements (underlined) that DO NOT COUNT as a generalisation because they are not built up in the lesson from examples:
    - A student asks why the teacher multiplies both sides of the equation by -1. The teacher replies that this is something students can do to make factorising easier.
    - The teacher is talking about substituting…T: 7. I have to replace to obtain those values.
    - Another example of a general statement:

      38:28 T: We are going to define…definition…

      38:33 SN: Sh.

      38:34 T: Write there as a title: quadratic…equation

      38:47 T: Definition.

      38:55 T: I don't know where my pen is, let's see.

      39:02 T: We define as…the equation…of second degree…or quadratic…equation…to the one…that has the form…two points…AX squared plus BX plus C equals 0.
- Evidence: teacher and students.

Table A B.18. Clarity

Component	1	2	3	4
Clarity. The extent to which the mathematical content around the learning goal of the lesson is presented clearly and students appear to follow along with the content of the lesson.	The mathematical concepts, tasks, student response patterns or discussions in the lesson are generally murky.	The mathematical concepts, tasks, student response patterns or discussions in the lesson have more murkiness than clarity.	The mathematical concepts, tasks, student response patterns or discussions in the lesson have more clarity than murkiness.	The mathematical concepts, tasks, student response patterns or discussions in the lesson are clear.
	There are multiple instances in which students demonstrate they do not understand the same logical element(s) of the lesson. There is a pattern to students' behaviours around clarity.	There are at least two instances in which students demonstrate they do not understand the same logical element(s) of the lesson. There is a pattern to students' behaviours around clarity.	There may be instances in which students demonstrate they do not understand the same logical element(s) of the lesson. There is a pattern to students' behaviours around clarity.	There are no instances in which students demonstrate they do not understand the same logical element(s) of the lesson. There is a pattern to students' behaviours around clarity

Definition of clarity

91. Logical elements: the component parts of a lesson's system reasoning. For example, a teacher's introduction to the topic, the procedures for solving a problem, and the examples the teacher reviews.

Notes and rating guidance

- Evidence to pay attention to when judging clarity:
  - logic is linear
  - ease of following steps (e.g. does not skip important steps)
  - clearly represented physically on worksheet or board
  - students' questions do not show a pattern related to the logic of the lesson
  - there are no major errors or there are not many minor errors.
- One way to determine if the mathematics is clearly presented is to observe students' questioning pattern. Questions do not necessarily mean students are confused – they can be interested or simply curious. But when multiple students ask the same question about how two things are related, or what the next step in the process is when the teacher has just said how they are related and what the steps are, this can be an indication of a lack of clarity. The rater should pay attention to both the presentation of the mathematics and students' responses to that presentation.
- The rater should ask: Can I follow the mathematics easily? Then the rater should review the evidence of clarity.
- Examples of students showing confusion:
  - Students keep asking the teacher similar clarifying questions when the teacher is walking around during small group work.
  - Students ask other students for clarification.

- Students make statements such as “I don’t understand”, “This doesn’t make sense”, “I think I am lost”; disengage in the assigned work, or make facial expressions suggesting they are confused.

- Rating guidance:
  - When the rater considers the murkiness or clarity of the evidence, if it was more clear, the rating will be a 3 or 4. If it was more murky, it will be a 1 or 2.
  - The rater should use behavioural evidence to arrive at a final rating.
  - Mathematical errors may make the lesson less clear to the students. To determine if the errors should lower the first descriptor rating you assign in your mind, consider whether the error **substantially impacts the students’ ability to accomplish the learning goal**. If the error does not meet this threshold, you do not need to lower the first descriptor’s rating in your mind.
- Evidence: teacher and students.

Table A B.19. Explicit learning goals

Indicator	1 Little explicitness	2 Some explicitness	3 Predominantly explicit
Explicit learning goals. The extent to which the teacher poses explicit learning goal(s) to students for the lesson and activities. Note: Ratings can be carried over from the previous segment if students are continuing with the same activity.	The teacher does not explicitly state or write the learning goal(s) or activities.	The teacher explicitly states or writes the activities or topics in which students will engage. There is no explicit statement of the learning goal(s).	The teacher explicitly states or writes the learning goal(s).

Definition of explicit learning goals

92. Verbal or written statements about what students are expected to learn in the lesson are learning goals stated explicitly.

Notes and rating guidance

- There may be more than one learning goal in a lesson. When the learning goal changes, the rater should code the new learning goal. Otherwise the code for learning goal should be carried over from previous segments.
- It is important for the rater to pay attention to exactly what the teacher says in order to code this properly.
- When lesson goals are written on the board, raters should follow the same logic described above. In addition, the following rule should apply:
  - If the teacher mentions the EXPLICIT goal written on the board it should be considered explicit.
  - If the teacher does not mention the EXPLICIT goal written on the board it should be considered implicit.
- The rater should ask herself whether the following statements can be completed to assist in determining the learning goal and its implicit or explicit nature.

  - The student will be able to …...
  - The student will understand/know……
- The rater can ask herself "Did the teacher name an activity the students will do or something they will learn?"
- Examples:
  - Implicit examples: "So yesterday we left off solving incomplete quadratic equations. Today we will start with the next section in the chapter." "We will start with reviewing your homework and move on to the new material after that."
  - Explicit examples: "Today we are going to work on understanding the relationships among three types of quadratic equations." "Today you are going to learn how to solve quadratic equations in two ways."
- How to rate naming an activity, giving directions, or naming topics:
  - Naming an activity: When the teacher names and activity the students will be doing, that should be counted as a 2. E.g. "Today we will go over incomplete and complete quadratic equations." "Today we are going to review everything we've learned on quadratic equations." "We will begin with a real-world problem and then move to general ways to solve quadratic equations."
  - Giving students directions: Directions for completing a problem do not count as evidence for a 2. The teacher must explicitly name the activity the students will complete at the grain size of the activity (not how to do a specific task). The learning goal code is designed to capture a larger grain size that orients students to what they are learning over a whole lesson, not a single problem. Raters should look at the directions on worksheets in front of students.
- Example of task directions: On the worksheet in front of the students, the worksheet reads "Below are 2 shapes, the larger shape has area A the smaller shape has area B. Write expressions for the shaded area of each of the shapes below." The teacher then says, "What I'd like you to come up with for me please are expressions that describe the total shape in each question. So what combination of A and B make up each of the following shapes?" What the teacher has said is a rephrasing of the directions for the problem.
  - Naming a topic: When teachers name a topic the students will be working on, this should be coded as a 2. E.g. "So the topic for today is quadratic equations."; "We'll continue today with completing the square."; "This lesson builds on yesterday's lesson. We will be looking at the p-q formula."
- Naming the topic that will be covered or stating and activity the students will do count as a 2.
- Evidence: teacher.

Table A B.20. Accuracy

Indicator	1 Significantly inaccurate	2 Somewhat inaccurate	3 Predominately inaccurate
Accuracy. The extent to which the public mathematics of the lesson is factually correct and accurate. There are no errors or imprecisions in the teacher's mathematics	Minor mathematical errors or imprecisions in the mathematics are present more than once OR there is one major error.	The mathematics of the lesson has at least one minor error or imprecision.	The mathematics of the lesson is correct and accurate.
	If imprecisions or inaccuracies exist, they are not corrected.	If there are imprecisions or inaccuracies, they are not corrected or are not corrected consistently.	If there are imprecisions or inaccuracies, they are corrected consistently.

Definition of accuracy

- Major error: Mathematical mistakes of various types that are built upon in subsequent work and/or significantly undermine students' ability to meet the learning goal. Major errors may include incorrect use of mathematical terms or representations, incorrectly solving problems, or incorrectly specifying a problem or the steps to solve it.
- Minor error or imprecision: Mathematical mistakes of various types that are not built upon in subsequent work. In general minor errors do not significantly impact students' ability to meet the learning goal. These mistakes may be incorrect conventions (e.g. incorrect use of symbols or axes), incorrect use of mathematical terms (e.g. incomplete and complete quadratic equations, equation, etc.) or incorrect use of representations (e.g. parabolas, tables, etc.).
    - Errors and imprecisions are not identical, but they are treated the same in the rating scheme, so it is not generally necessary to distinguish between them.

Notes and rating guidance

- Judgments are to be made based on the public content the teacher presents.
- If the rater cannot see the students' mathematical work, presume it is accurate unless there is evidence to the contrary.
- A corrected error should not be counted as an error.
    - If an error made in a previous segment was corrected, the rater should return to all previous segments where the error was counted and change the score to a 3.
- Raters need to have clear evidence to demonstrate that there is an error or imprecision. If the mathematics feels generally murky, that should be coded in Clarity (a component, not an indicator).
- It is most important that errors get resolved or addressed. If an error gets resolved in a later segment, raters can go back to change the accuracy rating in the previous segment(s).
- Students may make errors or have significant imprecisions. If they are public and the teacher leaves them uncorrected or unspecified, the rater should take this into account.
- Guidance on major and minor errors:
    - arithmetic errors – count as minor, as long as not built on

  - mis-speaking – count as minor, as long as not built on
  - incomplete specification -- count as minor
  - poorly specified problems – count as minor.
- Examples:
  - Teacher said discard the "x" in the equation, this is an example of imprecision; this is minor.
  - Teacher did not work the equation to 0; this is a minor error.
  - The teacher is helping students learn to solve word problems that include quadratic equations. The teacher does not tell the students to take account of the units in setting up the problem. As students work through the problem, the students and teacher realise that the units have not been properly converted. The class corrects the units and completes the problem. This is a major error because although they units eventually get corrected (the correction of arithmetic – and therefore, minor – error), the error of not teaching the students to take account of the problem's units when solving word problems is not corrected. The latter is a major error.
- Sometimes teachers introduce an error for a specific reason. Such a teaching decision might not become visible in the 8-minute segment. If this occurs, the rater should go back and change the rating from previous segment(s) to reflect this.
- Imprecisions/errors that we are choosing not to count because we do not know whether a number system has been specified.
  - Rationale: It is fairly common in our videos that a teacher and the class are clearly working within the real number system but there is no clear specification that this is the case. This may be because it is, in fact, implicit. Or it may be that it was specified in a prior class, or at the beginning of the unit. Because we don't have that information, we can't easily determine whether certain statements are correct or not. For example, a statement like "the equation has no roots" is generally not true, but might be true under the real number system.
  - Scoring rule: If there is no contextual information to tell us whether the students know what number system they are working in AND a statement might be true or false depending on what number system they are working in, we do NOT count it as a major or minor error or imprecision.
    - Example: The square root of negative numbers does not exist.
      - Teacher says that the following cannot be solved: x2 + 25 = 0 because "you cannot take the square root of a negative number".
      - T: If instead of having X squared minus sixteen I have X squared plus sixteen, what happens? It's solved in the same way. The X squared is left alone, I move sixteen to the other side, changing the sign. What happens here when I take the square root? What am I left with? The square root of a negative number. What happens with the square root of a negative number? This solution, therefore, doesn't exist… That is, this equation has no solution.
    - Example: We do not know if real-numbers coefficients have been specified. Complex roots are mentioned as something to come, but we don't know if this will be addressed much later (like later in high school) or tomorrow.

   - T: a, b, and c belong to the real numbers, OK? They belong to the real numbers. Let's move on. How many solutions will a quadratic equation have? Two. Those solutions, can they belong to the real numbers or another set? They could be complex numbers. So the coefficients belong to the real numbers, but the solutions may be real or complex numbers, which is something we will discuss later.

- Imprecisions/errors that we are choosing not to count because of between-country differences in what is considered accurate.
  - Rationale: On certain subtle points there may be differences from one country to another in what conventions are considered acceptable in language. In a few identified cases, raters are instructed to not rate these items as errors so that we can maintain the ability to make comparisons in the analysis. These cases are listed below.
  - Scoring rule: If a teacher (or student) states that a number has two square roots, one positive and one negative, do <u>not</u> consider this an error. Likewise, if a teacher (or student) states that a number has one square root and that the positive or negative of that root square to the number do <u>not</u> consider this an error. Different conventions may be in effect as to whether "a square root" refers to the principle root or may refer to any root.
    - Example: "So when we calculate the square root of 4 we get plus or minus two."
      - Caution: The language should otherwise be correct, so if the teacher overgeneralises and says a quadratic equation will always have a positive and negative root, that is still incorrect.
      - Caution: The square root symbol, absent a +/- or - sign in front of it, does conventionally denote the positive square root. Therefore a statement that that symbol always means both roots would still be considered incorrect. (The +/- generally needs to be written explicitly if that is what is meant.)
  - Scoring rule: Number of solutions to a quadratic equation. Do not count as an automatic error if a teacher (or student) states that the number of solutions to a quadratic is always 2. Likewise, do not count as an automatic error of the teacher states that the number of solutions can be 0, 1, or 2. Alternately, a teacher may describe a root as a single solution or as two solutions that are the same; we count neither statement as an error, because the convention around how this is described differs between countries.
    - Example: "T: And here, we also have two solutions that also belong to the real numbers. But if you notice, here the guys, Bryan's group got minus 2 and minus 2, right? They got two solutions that are equal, which is the same as having one solution, yes? They only got one solution.
    - Example: "And in general, in general there should be always two solutions to a quadratic equation. But sometimes they might be equivalent, or if we are only looking for real solutions there be no real solutions. "
      - Caution: Of course the description should still be accurate to the problem at hand, so if a teacher describes an equation with one solution as having none this would still be incorrect.
- Evidence: teacher and students.

Table A B.21. Real-world connections

Indicator	1 Little or no connection	2 Some connection	3 Strong evidence of connection
Real-world connections. The extent to which what is being learned is connected or applied to something outside of school – a real-life problem or a student's life experiences.	There are no connections or weak connection(s) between the mathematical content being learned and real-life problems or students' life experiences.	There is at least one moderate connection between the mathematical content being learned and real-life problems or students' life experiences.	There is more than one moderate connection or at least one strong connection between the mathematical content being learned and real-life problems or students' life experiences.

Definition of real-world connections

93. Real-world connections are connections between the mathematics and students' lives outside of school.

- Definitions and examples for weak, moderate, and strong connections
  - Weak: The connection does not contain details about either the mathematics or the real-world context. The link between the mathematics and/or the real-world context is vague.
  - Moderate: The connection contains some details about either the mathematics or the real-world context. The link between the mathematics and/or the real-world context is clear or vague.
  - Strong: The connection contains details about both the mathematics and the real-world context, linking these details to one another clearly.

Notes and rating guidance

- In order to count as a real-world connection, the connection must be related to the purpose of the lesson.
- Referencing something that occurs in students' experiences outside of school without connecting that to the mathematics to be learnt does not count as a connection.
- Rating rules:
  - Many weak connections should be rated a 1.
  - A mixture of moderate and weak connections should be rated a 2.
  - If the detailed mathematics being linked to concerns arithmetic, this should be considered a moderate connection. For example:

    S: We plug in negative 2, and then it would be negative 7 and then negative 1.- detailed.

    T: Great, and so, you're right.

    S: But if you multiply them together, it would equal 7.

    T: It would equal positive 7. But here is the thing; in the context of the problem, does it make sense to me to say: "Hey, JE, I need a door and the dimensions of the door are gonna be negative 7 meters by negative 1 meter". Does that make sense?

S: No.

T: No, right? But you're right that's it a correct answer, but for answering a world problem that is talking about area of a rectangle, it doesn't make sense, okay?

- Evidence: teacher and students.

Table A B.22. Connecting mathematical topics

Indicator	1 Little or no connection	2 Some connection	3 Strong evidence of connection
Connecting mathematical topics. The extent to which the topic being learned is connected to other mathematical topics.	There are no connections or weak connection(s) between the mathematical content being learned and other mathematical topics.	There is at least one moderate connection between the mathematical content being learned and other mathematical topics.	There is more than one moderate connection or at least one strong connection between the mathematical content being learned and other mathematical topics.

Definition of connecting mathematical topics

94. The extent to which the quadratic equations topic under consideration in the lesson is being learned is explicitly connected to topics outside of the quadratic equations subtopics measured in TALIS Video

Notes and rating guidance

- Completing the square and quadratic functions should NOT be counted in this code because they are a part of some jurisdiction's teaching of quadratic equations in the TALIS Video study.
- These topics have been defined in other TALIS Video instrumentation as being a part of quadratic equations and therefore if connections in the lesson are being made to these topics, they should NOT be counted as connections for this code.
  - Handling algebraic expressions (working with brackets and terms)
  - Binominal formulae: a2-b2 or a2+2ab+b2
  - Introducing one form of a quadratic equation
  - Solving quadratic equations by ...
    - completing the square
    - factorising
    - quadratic formula X= (-b +/- SQRT(b2-4ac)) / 2a
    - finding roots in a graphical representation
  - Discuss different cases of ax2+bx+c=0 depending on values of a, b, c
  - Quadratic functions (definition, plotting and transforming graphs, etc.)
    - Connections to algebraic processes that have to be carried out in order to solve a quadratic equation <u>do not count</u> as connections. Mathematical identities and rules also <u>do not count</u> as connections, no matter how formally named. For example, the following does NOT count as a connection:

    - Example 1: T: Bring it to the same denominator. 6th grade. You know this.
    - Example 2: T: The product of two numbers is 0, at least one of them is 0.
  - This is an example of a weak connection:
    - T: In the 8th grade we were only able to solve linear equations. If we had a squared variable there then we were hoping that somehow the squares would cancel down or dissolve. But now we have learned ...how easily we can solve a quadratic equation.
  - Examples of connecting mathematical topics:
    - Moderate – specific but somewhat vague
      - T: The same way we did with linear equations, where the objective was to isolate the unknown to find the value in it, we have to do something similar in this kind of equation. But we have to take into account that the way to solve these exercises will be different, since the variable in this case is quadratic.
    - Moderate—specific but somewhat vague
      - S: If I do that step then the problem of solving a quadratic equation with one unknown changes into the question of solving linear equation.
    - Strong – specific and detailed
      - T: How many solutions did you have in seventh grade? One, right? For simultaneous equations we did in eighth grade, how many pairs did you have? Only one pair, right? x is this and y is that. This is a new type of equation which gives two solutions. We're seeing this for the first time.
- Evidence: teacher and students.

Table A B.23. Mathematical summary

Indicator	1 Weak summary	2 Moderate summary	3 Strong summary
Mathematical summary. The extent to which the teacher or students provide a summary of the mathematics under consideration in this lesson. A summary is a review of what has or should have been learned in this lesson. Note: If there is more than 1 summary, rate the best one.	There is no summary of the mathematics being learned or the summary is implicit and/or vague.	There is at least one explicit summary of the mathematics being learned. The summary reviews a small amount of mathematical work the class has done. It may be clear or somewhat vague.	There is at least one explicit summary of the mathematics being learned. The summary reviews a significant amount of mathematical work the class has done. The summary

Definition of mathematical summary

- Explicit: Stated clearly, verbally or in written form.
- Implicit: Implied, not specifically or clearly stated.

Notes and rating guidance

- Summaries can be made by the teacher or the students.
- The only summaries that should be counted in this code are ones that review the mathematics in this lesson.
- Do not count summaries of previous or future lessons.
- If there are multiple summaries, use the best one and rate based on the quality of that summary.
- Code the summary in the segment in which it occurs.
- Summaries that are in front of the whole class OR to a smaller number of students (individuals, pairs, groups of students) should be counted.
- Example of mathematical summary:

  T: Let’s make a conclusion. What is the discriminant of Δ?

  E: b squared minus 4ac.

  T：So the condition of roots of quadratic equation with one unknown and the discriminant of roots? What’s the relationship between them?

  E: When b squared minus 4ac is greater than 0, the equation has two unequal real roots. When b squared minus 4ac is equal to 0, the equation has two equal real roots. When b squared minus 4ac is less than 0, the equation has no real roots.
- In order to move from one to two, the summary needs to include an explicit review of the mathematics in the lesson.
- If the summary reviews a significant amount of work, but is vague, rate it a 2.
- Evidence: teacher and students.

Table A B.24. Types of representation

Indicator	For each segment of time, code the type of representation used at any point in the segment. Representation(s) can occur as part of a problem or part of a solution. They can be used by students or the teacher. If one is present, record it.	
Types of representation. The type of representation used at any point in the segment.	1. Graphs (bar graphs or line graphs, for example)	a. 1 – not present b. 2 – present
	2. Tables (a table is an arrangement of numbers, signs, or words that exhibits a set of facts or relations in a definite, compact, and comprehensive form)	a. 1 – not present b. 2 – present
	3. Drawings or Diagrams (Drawing must include information relevant for solving the problem. It does not count as a drawing if the symbols are spatially arranged to highlight certain features, if arrows are pointing to certain symbols to highlight them, or if arrows or other nonstandard marks are used in place of standard symbols (e.g., an arrow is used instead of an equal sign)).	a. 1 – not present b. 2 – present
	4. Equations (e.g. y=ax2 + bx + c) and Expressions (e.g., 2x2+ 3x)	a. 1 – not present b. 2 – present
	5. Objects (physical objects, e.g. a sheet of paper; a miniature of the Eiffel Tower or San Francisco Golden Gate Bridge).	a. 1 – not present b. 2 – present

Definition of types of representation

- Graphs: Examples include, but are not limited to, bar graphs, line graphs, and graphs on the Cartesian plane.
- Tables: A table is an arrangement of numbers, signs, or words that exhibits a set of facts or relations in a definite, compact, and comprehensive form. Lists of ordered pairs count as a table.
- Drawings or Diagrams: Drawing must include information relevant for solving the problem. It does not count as a drawing if the symbols are spatially arranged to highlight certain features, if arrows are pointing to certain symbols to highlight them, or if arrows or other nonstandard marks are used in place of standard symbols (e.g. an arrow is used instead of an equal sign). A scheme or conceptual map would count as a drawing.
- Equations and expressions: Examples include y=ax2 + bx + c. Expressions include 2x2 + 4
- Objects: Physical objects used as representations, e.g. a sheet of paper; a miniature of the Eiffel Tower or San Francisco Golden Gate Bridge to show parabolas.

Notes and rating guidance

- The rater should not make judgments about how well a particular representation was used. This code is designed to capture the fact that the representation was present in the segment.
- Rating rules:
  - If a teacher is using a physical object for something other than learning about the lesson’s goal, do not count it.
  - If a teacher references a physical object, but the physical object is not in the room (e.g. the Eiffel Tower), do not count it.

- If there is a representation in the view of the camera, but it is not used by teachers or students, do not count it.
- If the teacher or students use a representation in a segment, it can continue to be counted as long as it is still visible to students AND the rater has reason to believe it may be being used by the students.
- If the students switch activities and the representation is still visible, but no longer a part of the new activity, do not count that representation.
- If a teacher draws a parabola that is intended to show a specific quadratic equation or idea about a quadratic equation, this should be coded as a graph (even if the axes are not marked); it should not be coded as a drawing.

- Examples that do not count:
  - The teacher has many triangles that are meant to be cut apart to make a puzzle. On each edge of each triangle, there is a single quadratic equation in some form. Students solve the equations to then match up the edges of the triangles to form a shape. The triangles and the resulting shape do not count because they do not have to do with the lesson's goal which is factorising and solving quadratic equations.
- Evidence: teacher and students.

Table A B.25. Organisation of Procedural Instruction

Indicator	1 Little organisation	2 Adequate organisation	3 Predominantly organised
Organisation of procedural instruction. The degree of organisation, detail and correctness in the presentation of content when describing procedures or the steps of a procedure. Procedures are instructions for completing a mathematical algorithm or task (e.g., presenting new procedures, reviewing previously learned material, and doing the steps or processes used in the context of solving a problem/problems).	There are no procedures. Or The procedures presented are somewhat correct. Or The procedures presented are correct, but are not particularly organised AND lack detail (i.e., are vague).	The procedures presented are correct, but are not particularly organised OR not particularly detailed.	The procedures presented are correct, well organised and detailed.

Definition of organisation of procedural instruction

95. Procedures are instructions for completing a mathematical algorithm or task (e.g. the presentation of new procedures, review of previously learned material, and descriptions of procedures used in the context of solving a problem/ problems).

Notes and rating guidance

- A teacher may teach a process or procedure for completing a task without explicitly identifying the process as one that could be used across similar tasks. These procedures should count as procedural instruction.
- The rater should ask herself:
  - Is a procedure being taught? Name it.
  - How correct, detailed, and organised is it?

- Features of organisation and detail
  - Physical organisation
    - supports note-taking
    - legible
    - if the student only had the notes, would she be able to follow the mathematics?
  - Logical organisation
    - linear
    - "asides" and "loops" are noted for students
    - steps may be identified and used consistently when using the procedure
    - there may be explanations for why certain mathematical procedures and processes are allowed or are sensible.
  - How to think about the procedure
    - modelling metacognition "So the first step I ask myself is what kind of quadratic I have" or "Now at this point, I have to go back and check that the roots make sense."
    - identifying clear steps
    - identifying ways to check if the student has carried out the procedure correctly.
  - Use of examples
    - The language used to work through the example is consistent across examples.
    - Comparisons and contrasts between examples show students important aspects of the procedure or process.
- When students are performing or doing repetitive algorithms, and there is not a procedure being taught, it should be rated a 1.
- When there are errors during procedures or processes, the rater must consider how those errors impact the students' opportunity to meet the learning goal(s).
  - If there are errors that significantly impact students' opportunities to meet the learning goal, then the segment should be rated a 1.
  - If there are errors that do not significantly impact students' opportunities to meet the learning goal, then depending on the other evidence available, a score of 2 or 3 may be considered.
- Evidence: teacher and students.

Student cognitive engagement

Note for all components in this domain

96. This domain focuses on students' cognitive engagement. It is common to see teachers at the front of the room, explaining multi-step mathematical thinking of procedures. We can presume students are engaged in this mathematics if we have spoken

or written evidence that students are going through the mathematics step by step with the teacher. If students sit mostly silently through such explanation, we cannot assume students are cognitively engaged with what the teacher is explaining. The rater has to make a determination regarding whether the students are following the thinking using students' behavioural (written or spoken) evidence. Evidence of "following" will come from students spoken or written behaviours

Table A B.26. Engagement in Cognitively Demanding Subject Matter

Component	1	2	3	4
Engagement in cognitively demanding subject matter. Students regularly engage in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness.	Students do not engage in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness. Or There is a single brief engagement with such work, but it is done only by 1-2 students.	Students occasionally engage in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness.	Students sometimes engage in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness.	Students frequently engage in analyses, creation, or evaluation work that is cognitively rich and requires thoughtfulness.

Definition of engagement in cognitively demanding subject matter

97. Cognitively demanding subject matter is defined as work that engages students in analysis, creation, or evaluation, is cognitively rich, and requires thoughtfulness.

Notes and rating guidance

- The following are more detailed explanations of how these terms are operationalised in the rubric.
  - Analysis: Detailed examination or exploration of the features and relationships among mathematical procedures, processes, ideas, topics, etc.
  - Creation: Formulating or inventing a way to solve a problem or devising a way to solve a new problem or type of problems.
  - Evaluation: Determining the significance or conditions of a mathematical idea, topic, representation, or process.
  - Cognitively rich and requires thoughtfulness: Work that engages students' cognitive processes beyond recall, recitation, and the rote application of procedures. Such work frequently requires students to grapple with problems and ideas and the relationships among mathematical ideas, topics, representations, and processes.
- If the tasks under consideration provide the opportunity for analysis, creation, or evaluation AND you have evidence students are completing the task, you may count this evidence.
- In all activity structures (whole class, small group, etc.), you must have clear evidence the students (not just the teacher) are cognitively engaged.
- Pay careful attention to students' spoken and written behaviours and contributions to determine if you have clear evidence the students are "following" when the teacher is explaining to the whole group.

- It may be useful to consider the amount of time spent on the activity as you determine the rating.
- This code rates the degree to which a classroom of students is engaged in analysis, creation, and evaluation work that is cognitively rich; however, particular activity structures can present challenges to rating some segments. If for example, a single student solves a word problem that requires creation and analysis at the board but the rest of the students sit silently watching the board and they do not take notes, we have conflicting evidence at an extreme. For all but one student, there is no evidence to support a 2, 3 or 4. For the one student, there is evidence that can support a 2, 3 or 4. We do want to give "credit" for the evidence from that one student, but it must be weighed against the evidence from the rest of the class. Such a segment would likely get a 2.
- Evidence: students.

Table A B.27. Multiple Approaches to and Perspectives on Reasoning

Component	1	2	3	4
Multiple approaches to and perspectives on reasoning. Students use multiple solution strategies and/or reasoning approaches.	Students generally use a single procedure or reasoning approach to solve the problem or type of problem.	Students generally use a single procedure or reasoning approach to solve the problem or type of problem.	Students generally use a single procedure or reasoning approach to solve the problem or type of problem.	Students generally use two procedures or reasoning approaches to solve the problem or type of problem.
	Or			Or
	There is no evidence of how many approaches students are using.	There is a brief use of a second procedure or reasoning by approach at least one student.	At least one student uses a second procedure or reasoning approach in some depth.	Students use more than two procedures or reasoning approaches to solve the problem or type of problem in some depth.

Definition of multiple approaches to and perspectives on reasoning

98. This code focuses on multiple approaches students use to solve problems, not multiple solutions they come up with. There are two ways multiple approaches may appear: the teacher might have various students solve one type of quadratic equation using different approaches; OR the teacher might have students approach a single problem using different approaches. Both may count as long as there is more than one way to solve the equation.

Notes and rating guidance

- Pay careful attention to whether the students are using multiple approaches. Students (not just the teacher) must be using multiple approaches.
- Example of multiple procedural or reasoning approaches:

  T: Let's take a look at how this group did it. We have the same rectangle here. The total is 20 m, and thus the sum of one length and one width is 10 m. It should be one half. You can see it, right? The total is 20 m, and one length plus one width is the half of it. Then, this part is x, so 10 minus x. Have the other four groups thought of such an idea?

  S: We did it another way.

  T: (looking over students' shoulder) Ah! Yes, your group did it a different way. Let's have you put this and the other up on the board.

- Rating guidance:
  - If there is no evidence that multiple approaches were used, raters should rate the segment a 1.
  - The use of multiple approaches must occur in the same segment to be counted.
  - When counting strategies that are "used" in a segment, the rater may round up.
- Evidence: students.

Table A B.28. Understanding of Subject Matter Procedures and Processes

Component	1	2	3	4
Understanding of subject matter procedures and processes. Students engage in opportunities to understand the rationale(s) for subject matter procedures and processes i.e. students state the goals or properties of procedures and processes, state why a procedure or a solution is the way it is, or visually designate the elements or steps in a process or procedure.	Students do not engage in procedures or processes. Or When students engage with procedures or processes there is no evidence that they attend to the rationale for the procedures and processes.	When students engage with procedures or processes, they occasionally attend to the rationale for the procedures and processes.	When students engage with procedures or processes, they sometimes attend to the rationale for the procedures and processes.	When students engage with procedures or processes, they frequently attend to the rationale for the procedures and processes.

Definition of understanding of subject matter procedures and processes

99. This focuses on students' understanding of the rationale(s) underlying procedures and processes. This is visible in their spoken words and written work.

Notes and rating guidance

- The goal of the code is to capture whether students understand why or how a procedure works or what makes that procedure or process appropriate. This is different from students understanding what a procedure is (i.e. what steps make up the procedure).
- When students understand the rationale(s) underlying procedures and processes they are able to articulate the rationale(s) for the logic, appropriateness, and correctness of specific procedures and processes or individual steps of a problem.
- Teachers frequently support and/or lead students to articulate the rationales for procedures and processes. Evidence from these interactions should be considered; however, the rater should have clear evidence of students' thinking (i.e. their spoken or written work).
- Evidence of students' understandings may take various forms. Students may 1) ask questions about or state the goals or properties of procedures and processes, 2) ask questions about or state why a procedure or process is the way it is, and/or 3) visually designate the rationale for elements or steps in a process or procedure.
- The rater should ask:

1. Are students engaged with procedures and processes in the segment?
2. What is the evidence that students are doing this cognitive work? There must be spoken or written evidence the students have an understanding of the rationale(s) for procedures and processes.
3. What is the evidence that students are engaged in these types of activities? Students ask questions about or state the goals or properties of procedures and processes, ask questions about or state why a procedure or a solution is the way it is, or visually designate the rationale for elements or steps in a process or procedure.

- Examples of behaviours that should be considered by raters:
  - Asking questions about or stating the goals or properties of procedures or processes.
    - Students might explain why they should use a factorisation approach instead of a “completing the square” approach.
    - Students might compare and contrast why one solution process is more advantageous than another.
  - Asking questions about or stating why a procedure or process is the way it is.
    - Students might ask why a specific procedure does not work for a problem they are working on.
  - Visually designate (e.g. pointing, drawing, graphing, or using hands to show how parts of the problem are related or fit together) the rationale for elements or steps in a process or procedure.
    - Students might point to how a square is divided and represented in equation form, identifying the specific terms in the equation related to the square.
- Examples of common activities that do not count as evidence of students’ understanding the rationale(s) for procedures and processes:
  - Students silently listen to a teacher explaining the rationale for a particular approach to solving a problem.
  - Students write out the solution to a problem step-by-step.
  - Students recount the steps they took to solve a problems (e.g. “First I rearranged the equation into the general form. Then I simplified and combined terms. Then I solved for x”).
- Rating guidance:
  - If a segment is characterised by students not writing and/or speaking about the rationales for procedures and processes, the score of 1 should be given.
  - If students state the steps of a procedure or process but do not explain the rationale, those interactions do not count as evidence of understanding the rationale for procedures and processes.
- Evidence: students.

Table A B.29. Metacognition

Indicator	1 No metacognition	2 Modest metacognition	3 Developed metacognition
Metacognition. The teacher asks students to engage in metacognition by explicitly asking students to reflect on their own thinking.	Students are not asked to engage in metacognition.	Students are asked to engage in metacognition briefly and/or superficially.	Students are asked to engage in metacognition longer than briefly and/or in some depth.

Definition of metacognition

100. Metacognition is the act of thinking about or reflecting upon one's own thinking.

Notes and rating guidance

- Examples of metacognition:
  - The students have been solving quadratic equations using different approaches and the class is reviewing them on the board. The teacher says, "These are good solutions. Ximena, can you look at number 7 and think about why you did what you did? Why did you think it was a good idea to use the complete the square approach?"
  - Students are working in groups and the teacher circulates to look over their group work.

    T: How are you thinking you will approach the problem? What approaches have you already considered?

    S1: We thought about factorising but that seemed too hard because of the fractions.

    S2: And we thought about using the quadratic method because we can always use that.

    T: Ok, did you consider completing the square?

    S1: No.
- Evidence: teacher and students.

Table A B.30. Repetitive use opportunities

Indicator	1 No repetition	2 Some repetition	3 A lot of repetition
Repetitive use opportunities. Students engage in the repetitious use of a specific skill or procedure.	Students did not engage in the repetitive use of a specific skill or procedure	Students repetitively used a specific skill or procedure for less than half of the segment.	Students repetitively used a specific skill or procedure for half or more of the segment.

Definition of repetitive use opportunities

101. This code refers to the opportunities students have to practice a particular skill/procedure repeatedly.

Notes and rating guidance

- The underlying idea the rater should keep in mind is that students need the opportunity to carry out the same pattern of thinking multiple times in order to become proficient.
- The degree to which there are repetitive use opportunities must be judged against the skill/procedure they are learning.
- The skill/procedure must be tied to the learning goal.
- There must be 2 or more problems of the same "type" in order for the condition of "repetitive" to be met.
- The rater must keep track of the time in this code.
- The rater should gather the artefacts to judge this code when needed.
- The rater should ask what the student is practicing.
  - Students might be practicing picking one strategy (of a few strategies) to solve the problem.
  - Students might be practicing a single strategy.
- If teacher and students are reviewing answers to homework or an independent activity, to arrive at a rating, ask the question, "To what degree is there evidence students are carrying out the procedure/mathematics during the review?"
  - For example, if teacher and students are reviewing homework answers and steps to arrive at those answers, this does not count as repetitive use opportunities. Students must be doing the work themselves for it to count.
- If students appear to be working on a group of problems and there are more than 2 of a type, this may count as repetitive use. You do NOT need to have evidence of exactly how many problems the students completed.
- If students have been given a group of problems to work on and then the teacher goes through these problems together with students on the board, but only one problem is reviewed in for the entire segment, this is viewed as a continuation of the problem set and therefore can be counted as a 3.
- Evidence: students.

Table A B.31. Technology for understanding

Indicator	1	2	3	4
Technology for understanding. Students use technology for conceptual understanding.	Technology that requires electricity is not used.	Technology is used for communication purposes.	Technology is used primarily for communication purposes but is used once in a limited way to support conceptual understanding.	Technology is used exclusively for conceptual understanding OR for a mixture of communication and conceptual purposes. It is used more than once to support conceptual understanding in a limited way OR at least once in a complete way.

Definition of technology for understanding

102. Technology for understanding requires a tool that requires electricity.

Notes and rating guidance

- The focus of the code is on conceptual understanding. There are many ways teachers use technology that do not directly promote conceptual understanding. The most common way is by using overhead projectors, document visualisers, and power point slides to convey information. This use of technology is principally for communication and therefore, does not count as technology that is being used to support conceptual understanding.
- Technology can be used in a limited way either by there being a very short amount of time in the segment, thereby not allowing the rater to fully understand the contribution the technology is making to students' understanding. It can also be limited in that it focuses on a relatively minor aspect of the mathematics or a small part of a more major aspect of the mathematics.
- Technology is used in a complete way when the rater is able to fully understand the contribution the technology is making to students' understanding and the technology supports students' understandings of a more major aspect of the mathematics.
- The rater should ask: Is the technology being used as a chalkboard might be used? If yes, consider rating it the segment a 2.
- At the 3 level, technology is being used to support conceptual understanding. Conceptual understanding is supported when students are working on
  - understanding why subject matter procedures and processes are logical
  - analysis, creation, or evaluation work.
- Evidence: teacher and students.

Table A B.32. Classroom technology

Indicator	Technology definition: tool that requires electricity
Classroom technology. Check any technology that was used.	1. Overhead projector or visualiser/document camera 2. Smartboard or projector 3. Graphing Calculator 4. Non-graphing calculator 5. Computer/laptop 6. Television 7. Tablet 8. Cell phone 9. No technology that requires electricity is used

Definition of classroom technology

103. Classroom technology focuses on the technology the teacher uses for the lesson.

Notes and rating guidance

- If something uses batteries, this counts as using electricity.
- Only code technology that is used (not just present in the room).
- Rating rule: If you code #2: smartboard or projector, do NOT code computer
- If the teacher is using more than one type of technology in one segment, enter the types of technology in numerical order from least to greatest. For example, if during one segment a teacher uses (6) a television and then (2) a smartboard, raters should enter 2 first and then 6 for types of classroom technology.
- Evidence: teacher and students.

Table A B.33. Student technology

Indicator	Technology definition: Tool that requires electricity.
Student technology. Check any technology that students used individually, in pairs, or in small groups.	1. Graphing Calculator 2. Non-graphing calculator 3. Computer/laptop 4. Tablet 5. Cell phone 6. No technology that requires electricity is used

Definition of student technology

104. Student technology focuses on the technology students use that supports mathematical learning.

Notes and rating guidance

- If something uses batteries, this counts as using electricity.
- Only code technology that is used (not just present in the room).
- Rating rules:
  - When some types of calculators are being used, if there is no evidence that the teacher asks the students to use their calculators to graph, assume students are using non-graphing calculators.

  - If it is made verbally or visually explicit that students are using the calculator application on their cell phones, then raters can code calculator (3) or (4) – depending on the whether they are using it to graph or not.
  - If students are using their cell phones, but either it is unclear if students are using the calculator application or it is clear students are using a different application to support mathematical learning, then the rater should code cell phone (8).
  - If students are using their cell phones, but it is clear they are not being used for mathematical learning (e.g. they are texting a friend), then raters should not code cell phone (8).
- If students are using more than one type of technology in one segment, enter the types of technology in numerical order from least to greatest. For example, if during one segment students use (8) a cell phone and then (3) a graphing calculator, raters should enter 3 first and then 8 for types of classroom technology.
- Evidence: students.

Table A B.34. Software use for learning

Indicator	
Software use for learning. Whether instructional software is used to assist or support learning of the mathematical topic through simulations, instructional games, interactive graphing, etc.	1. Not present 2. Present

Definition of software use for learning

105. This code focuses on the software used for learning in the classroom.

Notes and rating guidance

- Any software in the lesson should be considered under this code.
- A ‘use for learning’ must go beyond communication purposes.
- The focus is on learning activities that directly support conceptual understanding of mathematics. Conceptual understanding is happening when students are working on
  - understanding why subject matter procedures and processes are logical
  - analysis, creation or evaluation work.
- Evidence: teacher and students.

Assessment of and responses to student understanding

Table A B.35. Eliciting student thinking

Component	1	2	3	4
Eliciting student thinking. Questions, prompts and tasks elicit detailed student responses (written or spoken).	There is no student thinking present.	There is a small amount of student thinking present. Questions, prompts and tasks result in perfunctory student contributions that only concern answers, procedures, or the steps necessary for solving a problem.	There is a moderate amount of student thinking present. Questions, prompts and tasks result in detailed student contributions concerning answers, procedures, and the steps necessary for solving a problem.	There is a lot of student thinking present. Questions, prompts and tasks result in a mixture of student contributions concerning answers, procedures, the steps necessary for solving a problem, ideas and concepts. Contributions may be detailed or perfunctory.

Definition of eliciting student thinking

106. Student thinking is any contribution students make to the lesson – written or spoken.

Notes and rating guidance

- There are various types of student thinking as specified in the rubric.
- Detailed contributions: Contributions that have sufficient detail about the mathematics being worked on, not just short answers that give the answer or define a term, for example. This code accounts for the level of specificity and information provided in the students' contributions.
- Detailed contributions tend to be longer than less detailed contributions.
- Detailed contributions are those that reveal students' thinking processes or rationales.
- Thinking processes can be revealed by students' step by step solving of processes.
- The rater should look carefully at artefacts to determine what the students are working on independently.
- The rater should consider the written work students do at their desks, but there must be clear evidence of the nature of that work.
  - For example, the teacher might have the students work on a problem for 3-5 minutes at their desks independently but then ask one student to share the steps the student used to solve the problem for another 2-3 minutes. If the student then shares the steps with the whole class, thereby providing evidence to the rater of the work the students were doing during the independent work time, this whole amount of time (both the time students used to carry out the work and the time spent sharing the step in front of the whole class) should be considered when rating this code.
- Examples of contributions that generally concern answers, procedures, or the steps necessary for solving a problem.

T: For X - 8 to be a factor, I need to put an operator, either plus or minus, along with a constant value. So if I've done that right, I should be able to do that. Jack, can I do that? Can I put a value in front of it and what is that?

S: Um, three.

T: And the operator?

S: Plus.

- Evidence: teacher and students.

Table A B.36. Teacher Feedback

Component	1	2	3	4
Teacher feedback. Teacher responds to students' thinking via feedback loops that are focused on <u>why 1) the students' thinking is correct or incorrect 2) ideas/procedures are the way they are.</u>	There is one or no feedback loops.	There are a couple of feedback loops.	There is some feedback loops.	There are frequent feedback loops.
Teacher and student exchanges address the mathematics in a complete manner.	Teacher and student exchanges address the mathematics in a generally limited manner.	Teacher and student exchanges address the mathematics in a generally limited manner.	Teacher and student exchanges address the mathematics in a mixture of manners – both limited and complete.	Teacher and student exchanges address the mathematics in a complete manner.

Definition of teacher feedback

- Feedback loops: a loop is a back and forth exchange between the teacher and students around why 1) the students' thinking is correct or incorrect or 2) ideas/procedures are the way they are.
- Complete feedback: responses to students' contributions that address the mathematics at hand in a detailed fashion.
- Limited feedback: responses to students' contributions that address the mathematics at hand in a perfunctory (although perhaps adequate) fashion.

Notes and rating guidance

- This component focuses on the teacher's responses to student thinking
- There are two aspects of feedback that are of focus in this component:
  - The back and forth exchanges between the teacher and students focusing on <u>why</u> the students' understandings are <u>correct or incorrect</u>.
  - The extent to which teacher and students' exchanges address the mathematics in a complete manner.
- Such exchanges could happen when the teacher is working through a procedure/process with a student(s), when the teacher is helping a student(s) to understand why a solution is the way it is, etc.
- Feedback loops are counted in the first descriptor. The second descriptor counts all types of feedback.

- The student doesn't need to be the same as long as the loops are focused on the same substantive mathematical issues
- Feedback happens in all activity formats – whole group, small groups, pairs, and individual work.
- Evidence: teacher and students.

Table A B.37. Aligning Instruction to Present Student Thinking

Component	1	2	3	4
Aligning instruction to present student thinking. The teacher uses students' contributions.	The teacher does not use students' contributions.	The teacher rarely uses students' contributions.	The teacher sometimes use students' contributions.	The teacher frequently uses students' contributions.
If students make errors or struggle mathematically, the teacher provides cues or hints to support student understanding.	If students make errors or struggle mathematically, the teacher does not provide cues or hints to support student understanding.	If students make errors or struggle mathematically, the teacher rarely provide cues or hints to support student understanding	If students make errors or struggle mathematically, the teacher sometimes provide cues or hints to support student understanding	If students make errors or struggle mathematically, the teacher frequently provides cues or hints to support student understanding

Definition of aligning instruction to present student thinking

- Cues and hints: a comment or question that is intended to move a student's or students' thinking forward and is said in response to evidence of student thinking, whether that thinking is correct or not.

Notes and rating guidance

- There are four types of evidence that count as using student contributions
    1. drawing attention to the contribution or features of the contribution
    2. asking a question in response to a student's question or contribution
    3. having students provide the next step in the procedure or process
    4. acknowledging patterns in student contributions.
- Aligning efforts in a whole group or small group instruction context and the "one-on-one" context should be counted.
- Examples of a teacher using a student's contributions:
    - The student offers an answer "x+5" or "set it equal to zero" and responds with a question such as "why?", "how did you get that?", or "are you sure?"
    - Students are working in groups and the teacher selects groups to present their work in front of the whole class.
    - A teacher is solving a problem at the board and asks "What should I do next?" A student responds and the teacher follows the directions the student just gave.
    - A student gives an incorrect answer or mis-specifies a procedure and the teacher says, "Ok, there is a mistake here. Did anyone catch it?"
    - A student gives an answer and the teacher says to another student "Is that correct?"
- Examples of a teacher providing cues and hints:

T: How do you judge?

T: Can you find out the key words?

T: Look at it again, here, look at this side.

T: Find out real root, which condition are there real root?

E: It is greater than 0, one kind of condition.

T: Anything else?

E: It is equal to 0.

- Evidence: teacher and students.

Annex C. Rater agreement metrics in quality control processes and main study rating

Annex C presents five tables that show the rater agreement levels in the main study and in the quality control processes used to monitor raters (certification, calibration, and validation). All raters were trained via a train-the-trainer approach, using standardised English training materials and subtitled videos. A detailed description of how all metrics are calculated as well as more detail for each jurisdiction will be reported in technical report of the TALIS Video Study.

Components rater agreement

Table A C.1. Average rater-to-rater agreement statistics for the main study double ratings over the component codes in the TALIS Video Study policy report

Jurisdiction	Mean percentage exact	Mean percentage adjacent	Mean QWK	Mean Rating ICC
Chile	52%	89%	0.22	0.23
Colombia	54%	91%	0.23	0.25
England (UK)	50%	90%	0.20	0.22
Germany	53%	91%	0.31	0.32
Japan	50%	86%	0.22	0.24
Madrid (Spain)	55%	88%	0.23	0.23
Mexico	53%	90%	0.28	0.28
Shanghai (China)	56%	92%	0.12	0.16

Table A C.2. Average rater-to-rater agreement statistics for the main study double ratings for the three (analytic) domains based on components

Jurisdiction	Domain[1]	Mean percentage exact	Mean percentage adjacent	Mean QWK	Mean rating ICC
Chile	Classroom management	61%	91%	0.21	0.21
	Social-emotional support	42%	93%	0.28	0.30
	Instruction	52%	88%	0.21	0.23
Colombia	Classroom management	71%	96%	0.22	0.23
	Social-emotional support	45%	96%	0.26	0.30
	Instruction	51%	89%	0.23	0.24
England (UK)	Classroom management	68%	97%	0.18	0.19
	Social-emotional support	58%	96%	0.44	0.42
	Instruction	44%	86%	0.15	0.19
Germany	Classroom management	68%	97%	0.32	0.33
	Social-emotional support	50%	96%	0.35	0.33
	Instruction	50%	88%	0.29	0.32
Japan	Classroom management	79%	97%	0.09	0.10
	Social-emotional support	48%	90%	0.19	0.24
	Instruction	43%	83%	0.26	0.28
Madrid (Spain)	Classroom management	73%	96%	0.22	0.23
	Social-emotional support	59%	91%	0.34	0.32
	Instruction	49%	85%	0.22	0.22
Mexico	Classroom management	63%	93%	0.25	0.24
	Social-emotional support	48%	94%	0.31	0.32
	Instruction	51%	88%	0.28	0.29
Shanghai (China)	Classroom management	87%	100%	-0.02	0.19
	Social-emotional support	64%	98%	0.15	0.17
	Instruction	47%	89%	0.14	0.15

Note: [1]Social-emotional support excludes risk-taking component and Instruction is the combination of Discourse, Quality of Subject Matter (excluding Clarity), Cognitive Engagement, and Assessment of and Responses to Student Understanding.

Table A C.3. Average rater agreement statistics with master rater for certification, calibration, and validation over the components reported in the TALIS Video study policy report

	N Raters	Certification[1]			Calibration			Validation		
		Mean percentage exact	Mean percentage adjacent	Mean QWK	Mean percentage exact	Mean percentage adjacent	Mean QWK	Mean percentage exact	Mean percentage adjacent	Mean QWK
Chile	25	62%	95%	0.25	49%	87%	0.04	50%	91%	0.29
Colombia[2]	22	59%	97%	0.25	46%	87%	0.02	57%	95%	0.30
England (UK)	9	59%	94%	0.34	57%	93%	0.25	52%	90%	0.22
Germany	14	63%	95%	0.23	58%	94%	0.34	53%	91%	0.21
Japan	7	57%	92%	0.18	53%	91%	0.29	51%	87%	0.25
Madrid (Spain)	10	60%	97%	0.21	56%	92%	0.18	48%	91%	0.16
Mexico	20	59%	94%	0.25	51%	91%	0.15	51%	89%	0.22
Shanghai (China)	12	56%	92%	0.21	51%	92%	-0.02	49%	91%	0.15

Note: [1]Certification results are only presented for the raters who participated in the main study rating to have results for a consistent rater pool across the three quality checks.
[2]Colombia had 23 raters rate main study videos but one rater only rated a single main study video so did not participate in calibration and validation. Thus, this rater is excluded from calibration results as well to have a consistent rater pool across all three quality checks.

Indicators rater agreement

Table A C.4. Average rater agreement statistics between raters for double ratings of main study videos for indicators

Indicator type	Number of indicators	Rater agreement statistic	Chile (N raters = 25)	Colombia (N raters = 26)	England (UK) (N raters = 10)	Germany (N raters = 11)	Japan (N raters = 7)	Madrid (Spain) (N raters = 11)	Mexico (N raters = 15)	Shanghai (China) (N raters = 11)
All	38	Mean percentage exact	91%	90%	88%	88%	89%	91%	89%	89%
		Mean percentage adjacent	98%	98%	99%	98%	98%	99%	98%	98%
		Mean QWK	0.43	0.48	0.44	0.56	0.45	0.52	0.49	0.34
		Mean rating ICC	0.45	0.48	0.47	0.55	0.48	0.52	0.49	0.41
Max score = 2	22	Mean percentage exact	97%	97%	95%	95%	98%	98%	95%	98%
		Mean percentage adjacent	100%	100%	100%	100%	100%	100%	100%	100%
		Mean QWK	0.38	0.47	0.44	0.59	0.45	0.50	0.47	0.27
		Mean rating ICC	0.41	0.47	0.47	0.57	0.50	0.52	0.48	0.41
Max score = 3	9	Mean percentage exact	80%	77%	78%	75%	74%	79%	77%	69%
		Mean percentage adjacent	96%	96%	97%	96%	96%	96%	94%	91%
		Mean QWK	0.34	0.33	0.41	0.37	0.36	0.38	0.38	0.23
		Mean rating ICC	0.35	0.33	0.42	0.38	0.36	0.38	0.38	0.25
Max score = 4	7	Mean percentage exact	84%	86%	77%	84%	79%	85%	83%	87%
		Mean percentage adjacent	96%	97%	96%	94%	97%	97%	95%	98%
		Mean QWK	0.69	0.71	0.49	0.70	0.57	0.74	0.70	0.63
		Mean rating ICC	0.69	0.71	0.50	0.70	0.58	0.74	0.70	0.61

Table A C.5. Average rater agreement statistics with master rater for certification, calibration, and validation over all indicators and by indicator rating scale

	Indicator type	Number of Indicators	Accuracy statistic	Chile (N raters = 25)	Colombia (N raters = 26)	England (UK) (N raters = 10)	Germany (N raters = 11)	Japan (N raters = 7)	Madrid (Spain) (N raters = 11)	Mexico (N raters = 15)	Shanghai (China) (N raters = 11)
Certification	All	25	Mean percentage exact	85%	79%	86%	85%	82%	89%	84%	83%
			Mean percentage adjacent	97%	95%	97%	97%	95%	98%	96%	97%
			Mean QWK	0.39	0.30	0.48	0.47	0.33	0.60	0.45	0.41
	Max score = 2	7	Mean percentage exact	95%	91%	98%	99%	95%	100%	98%	96%
			Mean percentage adjacent	100%	100%	100%	100%	100%	100%	100%	100%
			Mean QWK	0.24	0.14	0.57	0.63	0.45	1.00	0.43	0.32
	Max score = 3	9	Mean percentage exact	78%	69%	77%	74%	74%	81%	75%	74%
			Mean percentage adjacent	96%	92%	97%	96%	95%	98%	94%	96%
			Mean QWK	0.35	0.21	0.34	0.33	0.22	0.50	0.35	0.31
	Max score = 4	7	Mean percentage exact	79%	78%	84%	84%	76%	85%	80%	80%
			Mean percentage adjacent	95%	94%	97%	95%	91%	95%	95%	96%
			Mean QWK	0.43	0.43	0.51	0.45	0.31	0.48	0.47	0.46
	Max score = 9	2	Mean percentage exact	95%	87%	89%	93%	88%	99%	92%	85%
			Mean percentage adjacent	97%	91%	92%	97%	95%	99%	95%	96%
			Mean QWK	0.92	0.78	0.85	0.96	0.77	0.98	0.87	0.91
Calibration	All	38	Mean percentage exact	88%	88%	90%	90%	89%	91%	90%	91%
			Mean percentage adjacent	97%	98%	99%	98%	98%	99%	98%	99%
			Mean QWK	0.25	0.25	0.41	0.50	0.47	0.51	0.42	0.05
	Max score = 2	22	Mean percentage exact	98%	97%	99%	98%	99%	99%	98%	98%
			Mean percentage adjacent	100%	100%	100%	100%	100%	100%	100%	100%
			Mean QWK	0.39	0.45	0.85	0.56	0.53	0.58	0.53	0.16
	Max score = 3	9	Mean percentage exact	65%	64%	71%	72%	71%	76%	69%	73%
			Mean percentage adjacent	91%	94%	95%	95%	95%	96%	94%	95%
			Mean QWK	0.12	0.17	0.22	0.31	0.30	0.35	0.23	0.00
	Max score = 4	7	Mean percentage exact	88%	87%	88%	89%	81%	87%	91%	93%
			Mean percentage adjacent	97%	98%	99%	96%	96%	97%	98%	99%

			Mean QWK	0.28	0.20	0.36	0.67	0.59	0.63	0.61	0.00
Validation	All	38	Mean percentage exact	91%	91%	88%	89%	89%	91%	89%	90%
			Mean percentage adjacent	99%	98%	98%	98%	98%	99%	98%	99%
			Mean QWK[3]	0.48	0.35	0.37	0.48	0.50	0.49	0.35	0.39
	Max score = 2	22	Mean percentage exact	98%	98%	96%	96%	97%	99%	97%	99%
			Mean percentage adjacent	100%	100%	100%	100%	100%	100%	100%	100%
			Mean QWK	0.48	0.36	0.33	0.43	0.45	0.51	0.19	0.41
	Max score = 3	9	Mean percentage exact	79%	76%	76%	77%	73%	77%	76%	71%
			Mean percentage adjacent	98%	97%	97%	97%	95%	97%	95%	95%
			Mean QWK[3]	0.39	0.34	0.35	0.42	0.38	0.32	0.40	0.21
	Max score = 4	7	Mean percentage exact	86%	89%	76%	82%	85%	86%	81%	86%
			Mean percentage adjacent	97%	95%	95%	94%	97%	98%	94%	98%
			Mean QWK	0.58	0.34	0.46	0.71	0.72	0.71	0.49	0.62