Beiträge empirischer Musikpädagogik

b:em, Vol. 16
October 2024
ISSN: 2190-317
DOI: 10.62563/bem.v16i2

Towards modelling instructional quality for music classrooms: Exploring subject-specific application of the TALIS/GTI generic framework

Cecilia Björk[1,2]
Bernhard Hofmann[3]
Jens Knigge[4]
Gabriele Puffer[5]

[1] Universität für Musik und darstellende Kunst Wien
[2] Åbo Akademi University
[3] Universität Augsburg
[4] Nord University, Campus Levanger
[5] Hochschule für Musik und Theater München

Zusammenfassung

Künstlerische Fächer werden bisher in Large-Scale-Studien zur Unterrichtsqualität und bei der Entwicklung entsprechender Forschungsmethoden nicht berücksichtigt. Der Beitrag untersucht, ob sich ein Beobachtungssystem, das auf generischen theoretischen Modellen basiert, auch für musikpädagogische Forschung sinnvoll nutzen lassen könnte. Dazu führten wir explorative Analysen videografierter Musikstunden mit dem Videobeobachtungssystem aus dem dritten Zyklus der OECD-Studie "Teaching and Learning International Survey" (TALIS/GTI) durch. Die dortige Modellierung von Unterrichtsqualität erwies sich als hilfreich; essenzielle musikpädagogische Kriterien dürfen aber nicht aus dem Blickfeld geraten. Eine systematische Diskussion fachspezifischer Qualitätskriterien könnte unser Verständnis von schulischem Musikunterricht schärfen und musikpädagogische Forschung besser mit interdisziplinären bildungswissenschaftlichen Diskursen verknüpfen.

Schlagwörter:
Beobachtungscodes, generische Modelle, large-scale-Studien, Musikpädagogik, schulischer Musikunterricht, Unterrichtsqualität, Videographie

Summary

Arts subjects have so far been absent from large-scale studies on instructional quality and related methodological development. The purpose of this article is to explore whether an observation scheme that builds on generic theoretical models might be meaningfully employed in general music education research and practice. We evaluated observation codes proposed in the video observation system from the third cycle of OECD's Teaching and Learning International Survey (TALIS/GTI) by piloting their use in analysis of music lessons on video. We found that while the theoretical models are helpful, over-reliance on generic criteria for instructional quality may marginalise the criteria that are essential for general music education. Overall, we argue that a systematic discussion of subject-specific quality criteria can sharpen our understandings and better connect music education research with interdisciplinary discourse in educational science.

Keywords:
generic models, general music education, instructional quality, large-scale studies, observation codes, videography

1. Background

International large-scale studies of instructional quality in compulsory school systems have constituted a prominent part of education research since the 1990s. The rationale for such studies is the need to provide accessible high-quality data, contribute to educational theory, and produce knowledge that teachers, researchers and policy makers can draw on to improve global quality and equity in education.[1] For instance, a key challenge for teachers all over the world is to plan and conduct lessons that provide a learning environment in which students are stimulated and motivated to develop new knowledge and insights (Baumert et al., 2013). Teacher education programmes that support preservice teachers in developing the competencies they need to create good conditions for learning are considered a key component of excellent educational systems. Such programmes cannot be designed without concrete ideas of what may be considered "good" or "high-quality" teaching. It is therefore important to explore possible dimensions of teaching quality both in theory and in empirical research.

While some of the existing studies on instructional quality focus mainly on student achievement, there has also been a broad range of cross-national and cross-sectional efforts to model classroom teaching practices in order to develop observational measures of instructional quality and to make progress in instrument validation (Hernández-Torrano & Courtney, 2021; Lietz et al., 2017; Wagemaker, 2014). Large-scale studies in which exploration and analysis of teacher practices are part of the research design include TALIS (Ainley & Carstens, 2018), TIMMS (Köller et al., 2001; Mullis et al., 2020), and PIRLS (Hopfenbeck & Lenkeit, 2018). Some theoretical models developed for use in these and other studies aim to capture generic teaching quality (Helmke, 2014; Klieme et al., 2009; Krauss et al., 2020, Praetorius et al., 2018; Praetorius et al., 2020b), while others have focused on subject-specific criteria (Charalambous & Litke, 2018; Kleickmann et al., 2020; Möller, 2016). Recently, attempts have also been made to bridge the gap between generic and subject-specific approaches (Praetorius et al., 2020a).

Arts subjects have so far been absent from large-scale studies on instructional quality as well as from related methodological development. The topic of teaching quality has been discussed in the context of music teacher evaluation and assessment (e.g., Orzolek, 2019), but not with explicit regard to research on instructional quality modelling. Theoretically elaborated and empirically based models designed to capture instructional quality are currently missing from research on general muc education (Kranefeld, 2021). This is possibly due to a deep-rooted scepticism among arts educators about the appropriateness of such approaches in general and for the arts in particular, and to challenges associated with cultural generalisation, standardisation and measurability in music education (Juntunen, 2015; Kertz-Welzel, 2015; Orzolek, 2019; Zandén, 2018). One particular worry is that potentially biased criteria and outcomes might be understood as monopolies that would then be upheld as guidelines for relevance and resource allocation within different educational systems[2] as well as in local schools without considering plurality, cultural context, or the creative and changing nature of the arts.

Some studies have nevertheless attempted to explore dimensions of instructional quality in music education from specific perspectives such as selection and structure of musical activities (Kranefeld & Dücker, 2013) and opportunities for "aesthetic experience and practice" (Wallbaum, 2018a). Niessen (2010) follows a different approach: given the importance of in-depth learning in music education, she builds on a framework of three "basic dimensions of instructional quality" that were first elaborated in the context of mathematics and science education (Klieme et al., 2009). Kranefeld (2021) discusses subject-specific differentiations and additions to a generic synthesis framework elaborated by Praetorius et al. (2020a), thus connecting the theorisation with interdisciplinary discourses. In the coming years, she suggests, music education research should continue to engage with generic dimensions of instructional quality, although she points out that the concretisation, theoretical elaboration and operationalisation of suggested subject-specific dimensions such as "aesthetic activation" and "aesthetic experience" will present a particular challenge (Kranefeld, 2021, p. 10).

Kranefeld's specific caveat is indeed well chosen because in international music education literature, particularly within the North American discourse, the concept "aesthetic" has carried histo-

[1] https://ilsa-gateway.org/ilsa-in-education

[2] We thank Dr. James Humberstone for helping us articulate this point.

rical baggage from the prolonged debate about music education as aesthetic education (MEAE) versus praxialist music education (e.g., Alperson, 2010; Bowman, 2005; Vogt, 2017). In the contemporary German-language music education literature, the terms "aesthetic experience" and "aesthetic practice" (*ästhetische Erfahrung*, see Rolle, 1999; *ästhetische Praxis*, see Wallbaum, 2018a) are well established and more broadly construed than in certain critical praxialist representations (see Mantie, 2016). Both concepts draw on theorisation that encompasses modes of subjective and often nonverbal everyday "life-world" experience as well as imagination and interpretation (Heß, 2018; Seel, 1985, 2005). However, we recognise that there may be pitfalls and substantial challenges in using certain terminology for international purposes and we agree with Kranefeld (2021) that there is more conceptual work to be done. Furthermore, existing challenges with regard to music-specific dimensions and basic assumptions about instructional quality are of course not limited to possible misapprehensions connected to the praxialism debate (Vogt, 2003). What the example reveals is that given the heterogeneity of cultural expressions and the (inter)cultural pluralism in music education, it is particularly important to be systematic and thoughtful when attempting to elaborate subject-specific quality criteria for research (on interdisciplinary challenges, see also Praetorius et al., 2020a, pp. 431–432).

2. Aim

The purpose of this article is to explore whether an observation scheme that builds on generic theoretical models, aims to capture instructional quality across school subjects, and has already been applied to a specific subject (mathematics) might be meaningfully employed in general music education so that it could serve as a basis for both empirical research and reflective practice. By exploring criteria from an international project that operationalises generic models and discussing them in the light of our knowledge and experience of quality criteria from general music education,[3] we will probe the theoretical feasibility of instructional quality modelling that integrates dimensions that may be considered relevant across subjects while respecting the specificity of music as a school subject.

3. Theoretical frameworks for modelling instructional quality

3.1 Connecting teaching characteristics and learning gains

One of the most influential generic models of instructional quality (Figure 1) in German-speaking countries was derived from both theoretical and empirical work. Empirical findings from the international TIMSS[4] video study showed connections between selected teaching characteristics and students' learning gains in mathematics lessons.

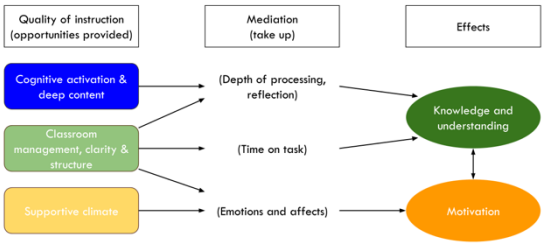

Figure 1. A theoretical model of basic (deep structure) dimensions of instructional quality and their effects on student learning and motivation (from Klieme et al., 2009, p. 140).

This model aims to enable description and evaluation of mathematics instruction, but it is also supposed to be generalisable across subjects and has already been used for conceptual and empirical work in several other school subjects, including general music education (e.g., Gebauer, 2013; Haas et al., 2019; Niessen, 2010). Building on both process-product and constructivist theories, teaching is understood as the offering of learning opportunities and is regarded as successful when students achieve a better conceptual understanding of the subject matter and feel strongly motivated to continue learning. Three dimensions of teaching are considered as relevant for students' learning gains:

1. Cognitive activation: The term refers to the quality of the learning tasks that are set by the teacher and thus to a "key dimension of the instructional quality of classroom learning" (Klieme et al., 2009, p. 140). Depending on the subject context, cognitive activation is characterised by "challenging tasks, activating prior knowledge, content-related discourse and participation practices" (ibid.).

2. High-quality, clear and disruption-preventive *classroom management* should provide students

[3] For details, see the section "Analytical approach".
[4] TIMSS = *Trends in International Mathematics and Science Study* (https://timss.bc.edu/).

with as much time as possible to focus on the learning task at hand (Klieme et al., 2001).

3. Student support: A supportive, student-oriented social climate is expected to produce positive motivational outcomes; theoretical grounding for this assumption is drawn from self-determination theory (Ryan & Deci, 2000).

However, it became apparent that this and other generic models of instructional quality were not suitable to adequately cover all relevant aspects for all school subjects (Praetorius et al., 2020a). As a result, subject-specific adaptations gradually gained importance. Building on several models from mathematics education as well as on concepts from educational research on other school subjects, Praetorius et al. (2020a) suggest an extended classroom observation framework that is intended to balance generic and subject-specific aspects of instructional quality (Figure 2):

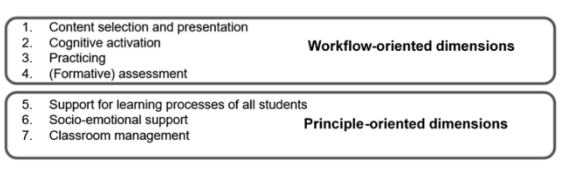

Figure 2. Seven basic dimensions of instructional quality in the classroom (see Praetorius & Charamboulos, 2018; Praetorius et al., 2020a; illustration by the authors).

The model is meant as a synthesis framework that emerged from an inductive, consensual approach in which shared and subject-specific dimensions, subdimensions, and observable indicators were taken into account. The authors identified a total of seven dimensions, including cognitive activation, classroom management, and supportive climate, and clustered them into two groups (Praetorius et al., 2020a, p. 412). The first four dimensions refer to the workflow of a prototypical teaching-learning sequence: (1) An appropriate selection and presentation of content which considers subject-specific methods and allows engagement with relevant learning objects; (2) Cognitive activation is intended to initiate and promote in-depth learning processes; (3) Well-supported practising helps to consolidate what has been learned; (4) Formative assessment is used to collect information for continuing and improving the learning process. The three further dimensions represent guiding principles for the overall design of the teaching-learning sequence: (5) Support for learning processes of all students refers to aspects such as differentiation and adaptivity; (6) Socio-emotional support is relevant for all steps of the learning process; (7) Effective classroom management and the resulting well-structured learning environment are fundamental prerequisites for successful learning.

Both models described above (Figure 1 and Figure 2) have been used as theoretical frameworks in international large-scale studies. One recent example is the TALIS/GTI study.

3.2 TALIS/GTI

In 2020, the Organisation for Economic Co-operation and Development (OECD, 2020) published the latest results from its study *Global Teaching InSights: A video study of teaching* (GTI). GTI is the third cycle of OECD's TALIS (*Teaching and Learning International Survey*) studies. Whereas the first two TALIS studies were conducted through teacher and principal surveys (asking about teaching and learning conditions), GTI is an international video study of teaching practices that looks "directly" into the classroom (OECD, 2017). GTI was designed to pilot new methods to capture real teaching practices and examine the analytical insights from generated data. GTI aimed 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 interrelated, and how contextual aspects of teaching are related to the student and teacher characteristics." (Opfer, 2020, p. 22)

In total, eight countries (Chile, Colombia, England, Germany, Japan, Spain, Mexico, China) participated in the study. The sample consisted of grade 8 and 9 students (N = 17,500) and their teachers (N = 700). All teachers had to teach mathematics, or more specifically, a focal unit of quadratic equations. These lessons were videotaped and then rated by trained observers. Various publications of results are now available (e.g., OECD, 2020; Zhu & Kaiser, 2022), mainly focusing on the connection between the quality of teaching and the competences of the pupils. While these results are undoubtedly intriguing and potentially motivating for similar study designs in the field of music education research, it is essential to acknowledge the necessity of conducting music-specific groundwork prior to further investigation. This is because in order to investigate the correlative relationships between teacher be-

haviour and student competence development, both dimensions of teaching must first be measured quantitatively. While various competence tests at the student level are available for this purpose (e.g., Hasselhorn & Knigge, 2021), there is still a lack of a valid assessment instrument for measuring the quality of instruction in music.

In light of the aforementioned considerations, our primary focus will be on two theoretical and methodological aspects of the GTI study: the theoretical model of the GTI study (3.2.1), and the methodological aspects of GTI (3.2.2), in particular the operationalisation of the theoretical model by means of an observation system and observation codes.

3.2.1 GTI's theoretical model

The conceptualisation of instructional quality in mathematics teaching that was used in GTI was derived by integrating three bodies of knowledge: participating jurisdictions' views of teaching quality, TALIS and PISA frameworks, and a review of international literature on teaching (Bell et al. 2020, p. 9; Opfer, 2020, p. 36). At a broad level, these multiple views were similar and could be synthesised in a model for capturing teaching quality.

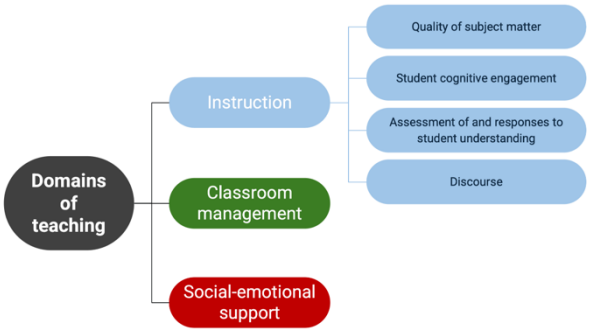

Figure 3. GTI's Domains of instructional quality (OECD, 2020; illustration by the authors).

On an overarching level, the GTI model contains three broad analytic domains of teaching (Figure 3): instruction, classroom management, and social-emotional support. The domain *Instruction* consists of four subdomains: *Quality of subject matter*, *Student cognitive engagement*, *Assessment of and responses to student understanding* and *Discourse*.

3.2.2 Operationalisation: one example

In order to specify the particular aspects of teaching that are being measured, each domain of analysis is further operationalised into *components* and *indicators*. A component is defined as "a code that applies to higher-inference classroom interactions" and is rated on a four-point scale (TVS International Consortium, 2019, p. 20). Indicators, on the other hand, are used to record "lower inference classroom events" and either count, categorise or rate the respective interactions on a three-point scale (ibid.).[5] Table 1 shows how the domain *Social-emotional support* is divided into individual indicators and components.

GTI's theoretical model relies on the assumption that "[g]roup learning of the type students experience in classrooms requires students to grapple with uncertainty" – and that "such processes require social-emotional support" (TVS International Consortium, 2019, p. 6; cf. OECD, 2019, p. 6). One crucial element that contributes to creating a supportive learning environment is the positive climate in the classroom, characterised by mutual respect, moments of encouragement, and shared warmth. Respectful language, positive tone of voice, and traditional manners are used in verbal communications (→ component *Respect*; Table 2), while smiling, laughter, joking, playfulness, enthusiasm, and verbal affection are observed in classroom interactions (→ component *Encouragement and warmth*; Bell et al., 2020, p. 11). Students' willingness to take risks (→ component *Risk-taking*) in the classroom serves as another indicator of social-emotional support. When students feel safe, they are more likely to ask questions, seek guidance from teachers or peers, volunteer to share ideas or opinions, or articulate their private thoughts (Bell et al., 2020, p. 11). Teachers may encourage this social-emotional risk-taking by asking students to share their private thinking with the whole class (→ component *Request for public sharing*; Bell et al., 2020, p. 11).

[5] The use of terms in this context deviates from the established conventions in educational research. Within the field, it is more common to use the terms "high-inference" and "low-inference" to refer to the observation or rating process itself (cf. Lotz et al., 2013), rather than the observed interactions. Low-inference ratings typically focus on directly observable behaviours requiring minimal interpretation, such as counting the number of questions a teacher asks during a lesson. Conversely, high-inference ratings always necessitate some degree of interpretation and judgement, often involving complex constructs that are not directly observable.

Table 1. Domain Social-emotional support: Components, holistic domain ratings, and indicators (OECD, 2018); illustration by the authors (cf. Björk et al., 2022).

Indicators*	Persistence	Request for public sharing		
Components and Holistic domain ratings**	Social-emotional support (overall)	Encouragement and warmth	Risk-taking	Respect
* Code that applies to lower inference classroom interactions; to be rated every 8 minutes)				
** Code that applies to higher inference classroom interactions; to be rated every 16 minutes)				

Learning requires intellectual and sometimes emotional challenges, which may manifest as errors or difficulties. It is critical for students to persist through these challenges to learn effectively (→ indicator *Persistence*; Bell et al., 2020, p. 11). According to GTI's model, teachers and students should demonstrate patience and encouragement, building an acceptance-oriented environment based on trust that makes students feel comfortable and secure enough to take risks when overcoming various challenges (Bell et al., 2020, p. 11).

Table 2 shows the last and most concrete level of operationalisation, the so-called "observation codes". For the sake of simplicity, only the observation codes for one single component, namely *Respect* (see Table 1), are listed here.

Table 2. Observation codes: Component Respect (TVS International Consortium, 2019, p. 32); emphasis in the original.

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 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.

The code scale ranges from 1 to 4, with 1 being the lowest and 4 the highest quality. The logic here is quantitative: the more respectful behaviour and the less disrespectful behaviour that can be observed, the higher the code that can be assigned. In addition to such observational code overviews, the TALIS video study observation system also contains a range of further "notes and rating guidance" for each component and indicator. These often include concrete (transcribed) examples of classroom interactions and how to code them, including instructions for determining the boundary between two categories and for judging frequency within defined time segments (TVS International Consortium, 2019, pp. 24–25).

Because space in this article is limited, we can only present a highly condensed summary of conceptual and methodological aspects of the Global Teaching InSights project, using a small number of examples. Overall, however, our work with GTI has given us the impression that this approach might also be interesting for music education research – in our opinion for (at least) three reasons. (1) In observation-based research on instructional quality, music education has so far lacked a validated set of research instruments. GTI has explicitly empha-

sised the piloting of methods, so that a variety of experiences and materials has been generated that can potentially be useful for research in music education. (2) These materials (from the theoretical framework to concrete observation codes including training material for raters) are documented so extensively and in such detail as is, to our knowledge, extremely rare in pedagogical research in general.[6] This results in concrete possibilities for application and further development that go far beyond the reception of study results. (3) Finally, the methods and materials mentioned in (1) and (2) provide a degree of transparency (especially regarding the operationalisation of theoretical assumptions/decisions) that can make subject-specific aspects visible and thus both enable an adaptation for another subject (music) and expose the drawbacks and limits of the approach.

4. Analytical approach

To explore the feasibility of adapting a generic model for analysing teaching in general music, we conducted a pilot study using the TALIS/GTI observation system and observation codes (TVS International Consortium, 2019) to analyse video material from an international research project on music education (Wallbaum, 2018b). All four authors have extensive experience of general music education as teachers, supervisors, and researchers, albeit (inevitably) within limited cultural contexts (Germany and the Nordic countries). In the first phase of the pilot study, we took a broad exploratory and heuristic approach, working both deductively and inductively. We probed all the domains, components and indicators using examples from the entire empirical material in order to compare quality criteria from TALIS/GTI with criteria from general music education literature (Kranefeld, 2021; e.g., *musikalische Vielfalt als Normalfall* (musical diversity as the norm), *ästhetische Aktivierung* (aesthetic activation), *senso-motorische Aktivierung* (sensorimotor activation); *Musikdidaktik* (e.g., Kaiser, 2010; Jank, 2021), *learning music through embodied and sociocultural processes* (Wiggins, 2015), *transformative music engagement* (O'Neill, 2012), etc.), from our own previous work in research and in curriculum and course development, and from our professional experience as teachers. We also engaged in critical examination of our own tacit beliefs as they emerged in contact with the empirical material and in our discussions.

We began with an exploratory application of the TALIS/GTI observation codes to several videos. We discussed goodness of fit and noted cases that seemed unproblematic as well as cases where we agreed that capturing music-specific instructional quality would require adaptation, differentiation, additions, or complete revision of the criteria. After this initial phase, we selected the two videos that corresponded most closely with our respective cultural, educational, and linguistic backgrounds: the "Bavaria-Lesson" and the "Sweden-Lesson", assuming that this familiarity would improve our chances of interpreting accurately what the teachers were striving for.[7] This aligns with a guiding principle of TALIS, according to which raters only evaluate lessons from their own cultural context (TVS International Consortium, 2019, p. 5).

We then conducted a more systematic analysis, focusing only on one of the six domains of teaching measured in the TALIS/GTI observation codes, namely *Social-emotional support* (Ainley & Carstens, 2018, p. 54; Bell et al., 2020). Unlike procedures that attempt to assess teaching characteristics on the basis of very short observations, such as the "thin-slice" technique (cf. Begrich et al., 2017), TALIS always analyses the entire lesson across six domains. The full TALIS/GTI framework was therefore too extensive and detailed to cover in an exploratory study. Among the domains, social-emotional support seemed most immediately relevant for general music and therefore most interesting for us to select. While the construct may be challenging to operationalise and observe, previous research suggests that the aspects described in the domain's components and indicators are indeed particularly important for subjects such as music education and sports (e.g., Herrmann & Gerlach, 2020; Kranefeld, 2021; Steinbach, 2018). Each of the authors rated the lessons independently, following the TALIS/GTI coding scheme and instructions (TVS International Consortium, 2019). The transcripts of verbal interactions provided by Wallbaum (2018b) proved to be a helpful basis for this step. We then compared and discussed our ratings, focusing on disagreements and on components/indicators left blank because at least one of us had not considered them applicable. During this work we came to the conclusion that most of the components and

[6] See in particular the comprehensive collection of material available here: https://www.oecd.org/en/about/projects/global-teaching-insights.html

[7] All videos from the project *Comparing international music lessons on video* are available at https://comparing.video.

indicators we rated were highly inferential items, the use of which requires careful thought and reflection. While the analysis was more detailed in this final stage, our purpose was not to measure interrater reliability, test specific item validity, or provide conclusive findings for the domain, but to explore and discuss the overall meaningfulness of the approach.

5. Findings

In this section, we present the results of our explorations and structure them in four categories. We did not expect the TALIS/GTI observation codes to provide a perfect fit for general music education, and during rating, all of us had, independently of one another, entered question marks instead of score points for several components and indicators (see "cases of c)", below). Overall, however, the TALIS/GTI system transferred reasonably well to music education for the domain of *Social-emotional support*. For instance, we were able to observe numerous instances of respectful listening and positive verbal or nonverbal encouragement. In the coding scheme, the component *Risk-taking* is defined as "the extent to which students are willing to share their thinking with the class voluntarily" by asking questions, asking for guidance, and sharing private work publicly, making "their internal thinking or problem-solving process available for their peers to read or hear" (TVS International Consortium, 2019, p. 35). Such public sharing of ongoing development is ubiquitous in music education where learning often takes place in real time and in front of peers and teachers who can hear and see both progress and mistakes. An encouraging atmosphere characterised by patience, shared warmth, verbal affection and benevolent humour provides support for persisting through challenges (TVS International Consortium, 2019, p. 7). Creating a supportive climate in the music classroom may include developing adaptive teaching strategies, establishing an environment that promotes experimenting and personal agency (see Odena, 2018; Wiggins & Espeland, 2018), and preventing situations that could cause embarrassment or shame, such as pushing students to perform in front of the class against their will. In the coding scheme for this domain, using musically relevant terminology, for example adding "music-making" or "creative processes" to "thinking", could allow music education researchers to reuse the component *Risk-taking* almost as it stands (Table 3).

Table 3. Observation codes: Component Risk-taking (TVS International Consortium, 2019, p. 35).

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.

Detailed coding work with the domain *Social-emotional support* allowed us to develop a critical, heuristic categorisation of TALIS/GTI descriptors (components and indicators for each of the six domains that are rated) in relation to music-specific criteria for instructional quality:

a) cases where a component or indicator and its descriptors transfer well and could be reused in music education research *as such or with only minor changes*,
b) cases where *differentiations and additions* seem necessary, and
c) cases where component, indicator or descriptors *do not seem appropriate at all* and we would not recommend using them in a music-specific model.

In addition, a fourth category would include

d) elements that are missing altogether from the generic model so that adaptation of existing elements is not sufficient and *entirely new components/indicators* (and possibly even domains) should be generated for a music-specific model.

Examples of category a) could include the entire domain *Social-emotional support* (as described above) and the domain *Classroom management*, described in TALIS/GTI as, for instance, the ability to handle routines in the classroom in a way that favours a focus on teaching and learning (van Tartwijk & Hammerness, 2011; TVS International Consortium, 2019, p. 6). In music lessons, routines include setting up a learning environment that

provides sufficient opportunity to engage in hands-on musical activities such as performing and creating music, plan suitable musical tasks, and monitor each student's progress. At the same time, there has to be a reasonable degree of order and the sound volume needs to be kept at a level that is not disruptive, which can be a particular challenge when various kinds of instruments and equipment are involved. Examples of category b) could include the indicator *Metacognition* within the domain *Student cognitive activation* (Table 4). The importance of developing metacognitive skills for musical learning and practising is well documented (e.g., Concina, 2019; Hallam, 2001) and is also described in a substantial literature on self-regulated musical learning (e.g., Godau, 2016; McPherson et al., 2017; McPherson & Renwick, 2001).

Table 4. Observation codes: Component Metacognition (TVS International Consortium, 2019, p. 68).

Component	1	2	3
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.

However, self-monitoring in musical learning is certainly not limited to thought processes. Learners may, for instance: listen closely to tone and intonation, remind themselves through internal self-talk of how to cope technically with difficult passages ("relax the shoulder and breathe in here"), strive to maintain rhythmic accuracy and stability, concentrate on shaping phrases, pay attention to nuances of emotion and subtle shifts in qualities of sound, draw together "connections among artistic and aesthetic ideas" (López-Íñiguez & McPherson, 2020), be aware of their emotions and self-expression, and respond with awareness and sensitivity to what those they are making music with are doing and expressing. All of these skills require meta-level active monitoring of self, others, and the musical flow. What needs to be engaged in, then, is not merely "work that is cognitively rich and requires thoughtfulness" (Table 5), but work that is musically rich and requires musical awareness and sensitivity. Furthermore, "analyses, creation, or evaluation" (Table 5) often take place nonverbally and in real time during music making. As Brandstätter (2008) points out, "[w]hile concept-bound cognition is primarily concerned with systematisation and the grasping of regularities, aesthetic cognition always returns again and again to the particularity of unique perception and experience. The continuous return to sensory experience means that the particularity of perception does not disappear behind the generality of the concept" (p. 103; translation by the authors).

For raters using a coding scheme, being able to identify evidence of those many forms of musical engagement is such a major challenge that it might in part disqualify even a reworked version of the component *Metacognition* as a case of b) (not inappropriate, but in need of differentiation or additions) in our heuristic categorisation. TALIS/GTI acknowledges corresponding challenges for mathematical thinking, pointing out that if students sit silently through an explanation given by the teacher, it cannot be automatically assumed that they are cognitively engaged and "following" the thinking. The rater can only rely on students' (written or spoken) "behavioural evidence" (TVS International Consortium, 2019, p. 5). Similarly, evidence of engagement in music might consist of visible and audible musical activities (e.g., singing, body percussion, improvising, moving to a beat, or musical self-talk such as humming while creating or analysing music). However, such physical behaviour is no guarantee that there is subjectively experienced affective, narrative, or social engagement with music (Greenberg & Rentfrow, 2015). Cognitive engagement is also uncertain: it is a methodological challenge to "make students' musical thinking visible and audible" (Björk et al., 2021). Straightforward examples of c) (component, indicator or descriptors do not seem appropriate) were harder to find as many of our candidates could be moved from c) to b) after discussion, adaptation and/or additions. For instance, the definition of "explanation" in GTI centres around descriptive statements that clarify, rationalise, or justify (TVS International Consortium, 2019, p. 41). Musical clarification, however, can also be accomplished through nonverbal "statements" such as fine-tuned gestures, expressive movement, or repeated, clear demonstration.

One indicator directly related to the subject matter in GTI, *Mathematical summary*, might be considered at least partly out of place for music education. The definition is "the extent to which the teacher or students provide a summary of the mathematics

under consideration in th[e] lesson. A summary is a review of what has or should have been learned in th[e] lesson" (TVS International Consortium, 2019, p. 59). Even in a reworked form, it would seem awkward to claim that a marker of quality in music education is that the teacher provides summaries, especially with the requirement that they should be neither "implicit" nor "vague" (p. 59). For declarative knowledge (e.g., the difference between a major and a minor third) such summaries can make sense, but in lessons designed to provide conditions for strong musical experiences, for instance (Gabrielsson, 2011; Ray, 2004), it seems unlikely that intense emotion or transformative change (Dewey, 1934; Westerlund, 2002) could be enhanced for educational purposes or even meaningfully described by a teacher's concluding and perhaps trivialising rationalisation.

Table 5. Observation codes: Component Engagement in cognitively demanding subject matter (TVS International Consortium, 2019, p. 64).

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.

Finally, for category d) (elements missing altogether), we often discussed whether some aspect of what we consider good music education could be added to and subsumed under the existing domain *Quality of subject matter* and therefore be considered an adaptation rather than something entirely new. For example, creative and interpretive activity are essential parts of musical learning and should in our opinion be considered fundamentals of subject matter and lesson content in music, whereas the space for artistic creativity is – unsurprisingly – limited in clearly defined units on quadratic equations (although creative reasoning and problem-solving are well-known dimensions of mathematical ingenuity in general). In the case of creativity and artistic interpretation, we still argue that their importance in and for music could even warrant a domain of their own if a similar model-based instrument were developed for music education. It is possible that the TALIS/GTI components and indicators, given the subject context in which they were developed, work best when learning outcomes (both cognitive and non-cognitive) are largely given and accepted in advance (see Uljens & Ylimäki, 2017, p. 11).

6. Discussion and outlook

The purpose of this article was to explore whether an observation scheme that builds on generic theoretical models aiming to capture instructional quality across school subjects might be meaningfully employed in general music education so that it could serve as a basis for empirical research and reflective practice. We found that the generic models are helpful and may have more to offer both research and practice in music education than we initially expected, and that indiscriminate scepticism towards these approaches is not justified. In addition, an adapted model could be useful when communicating with researchers from other disciplines or participating in interdisciplinary projects. Based on our experience with TALIS/GTI, however, we note that music-specific adaptation of generic models and observation schemes also poses a number of problems.

First, as a video study, TALIS/GTI focuses on observable domains. There are aspects of musical learning and experience that cannot be captured by observation only. One such aspect is emotion, arguably one of the most important dimensions of music (Krause & Oberhaus, 2012). In the TALIS/GTI

instructions and accounts, emotions appear to be taken into consideration only insofar as they are related to learning: "Group learning of the type students experience in classrooms requires students to grapple with uncertainty. Such processes require social-emotional support" (TVS International Consortium, 2019, p. 6). Second, the language used in the models, and therefore in the observation codes too, is primarily drawn from cognitive psychology and leans towards computational metaphors (Casey & Moran, 1989) such as *processing* and *activation* which are not typical of the language commonly used in the context of artistic/musical practice or to convey musical experience. Descriptions of subject-specific quality in music education may be better served by the kind of reporting that is characteristic of ethnography, narrative analysis, and other qualitative methods. Bautista et al. (2019) note that most empirical research on how classroom videos may support pre- and/or in-service school music teachers has indeed been qualitative; however, the authors remark that music teacher education has made little use of classroom video compared to other areas such as mathematics, science, and literacy teacher education, and that there is ample room for methodological development. As part of this development, and as Kranefeld (2021) has already argued, further subject-specific conceptual work is needed.

We did not find anything in the observation system that was completely irrelevant on a higher conceptual level. But even though some of the things a music teacher and a mathematics teacher do and strive for are the same, we think that the fit should not be overestimated. Just because we can agree with many of the quality criteria in the generic models, it does not mean that what is missing is unimportant. On the contrary, some of the subject-specific criteria that are absent represent precisely what is closest to our hearts as music educators because they are related to lessons "in music", not just "about music". Hence, the disparities between generic and specific quality can be visible in how a music educator would weight criteria: for instance, what might look like insufficient classroom management in mathematics may in some cases be precisely the kind of experimental, open and supportive environment that can favour musical learning. While some of the generic background is shared with other subjects, quality criteria become more "musical" the closer we come to actual music lessons.

There are additional methodological limitations. In TALIS/GTI, generalised models are operationalised through strictly controlled lesson content (quadratic equations in grade 8). The quality (or qualities) we are interested in regarding general music education have a broader scope. It is worth considering whether a more limited focus would give meaningful results. The purpose of TALIS/GTI is to identify differences within a highly standardised framework. However, quadratic equations are the same all over the world, whereas in music lessons, the subject matter itself varies considerably. The contents and purposes of the lessons available in the international video material we worked with also vary. This is the case in the two examples we chose, even though the age group is the same (13–15-year-olds). The "Bavaria-Lesson" from a German *Mittelschule* class focuses on atmosphere: "a configuration of different aspects (or points) of the classroom praxis, such as sounds, spatial relationships, postures, gestures, movements and communications" explored to a great extent through improvisation (Wallbaum, 2018a, p. 129), whereas the "Sweden-Lesson" from a lower secondary school features small-group instrumental learning and music-making in bands within the pop-rock tradition (Zandén, 2018).

A further issue concerns ecological validity: a single video recording only provides limited information about a teacher's work in the music classroom. The measures are confined to teachers' and students' observable behaviour during lessons, but they do not allow us to make any statements about learning outcomes in general or about short-term or long-term effects of the teaching and learning practices we have observed (short/formative vs. long/summative, see e.g., Klieme et al., 2009, p. 151).[8] The limitations of the TALIS/GTI study also apply to our efforts; for instance, the possibility that videotaping lessons affects student or teacher behaviour so that "typical" lessons are not captured (Opfer et al., 2020, p. 43). Music education also still largely lacks criteria such as those used in TIMMS and other studies where students' progress is recorded during a shorter (e.g., three consecutive lessons) or longer time (e.g., a school year). Moreover, some research suggests that different relevant aspects of "instructional quality" will emerge depending on the chosen perspective (e.g,. Baumert &

[8] For this purpose, TALIS uses a completely different set of data, namely the pupil results from the PISA assessments.

Kunter, 2006; Praetorius et al., 2018). TALIS assesses teaching quality from the perspective of teachers, teacher educators and researchers, but students' experiences are taken into account only indirectly via questionnaires and tests. The validity of the TALIS ratings is therefore limited to the observers' perspective, and we agree with Praetorius et al. (2018) that "[i]t is still unclear … how to deal with the fact that these perspectives often do not converge in their perceptions" [of how a teacher would weight criteria] (p. 423).

Last but not least, a critical discussion of normativity is necessary. All our observations and ratings are grounded in certain ideas about what "good" music education should be, and such ideas are inevitably connected with quality criteria drawn from different cultural and musical practices. The TALIS domain *Quality of subject matter* can perhaps stretch to include more criteria, but raters will then need to be familiar with cultural contexts and also with the musical genre(s) taught. Even in the TALIS/GTI study where the subject matter is essentially universal, raters were selected to observe only videos from their own school systems, and "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" (TVS International Consortium, 2019, p. 15). We did indeed experience instances of disagreement even between the four of us, given the cultural differences between the German and the Nordic music education traditions; this also revealed some blind spots and narrow perspectives of our own, and pointed to the value of joint reflection. In the description of the TALIS observation system, the authors explain that "[t]he codes steer the attention to specific aspects of teaching which are considered of higher quality by the global education community" (TVS International Consortium, 2019, p. 4). The extent to which consensus within such a "global community" can exist in the world of music education is not clear. Further studies might explore degrees of agreement about teaching quality in general music education across cultures that differ from each other far more than the German and Nordic traditions, an approach that could also promote intercultural learning.

In the research process we report on in this article, normativity was not the aim. We examined whether some criteria from generic models of instructional quality might be applicable in music education research and reflective practice (or at least adaptable), and we have argued that this is the case. The risk when attempting to apply quality criteria that are intended to be generic is that those aspects may be overemphasised while criteria that are essential for music education are pushed out of sight. However, the interesting points, as described above, are precisely those where generic criteria do *not* fit our subject. Such "friction points" which compel us as music educators to articulate the specifics of what we value may also provide incentives for research and fruitful discussion topics for reflective practice. Despite the challenges considered here, we conclude that modelling instructional quality in music classrooms is worthwhile in itself and that discussing quality criteria systematically can both sharpen our views and understandings of the subject and connect music education research more closely with interdisciplinary discourse, for example studies on other school subjects, and with a much-needed, broad dialogue within the educational sciences.

References

Ainley, J., & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018: Conceptual framework*. OECD Education Working Papers. Advance online publication. https://doi.org/10.1787/799337c2-en

Alperson, P. (2010). Robust praxialism and the anti-aesthetic turn. *Philosophy of Music Education Review, 18*, 171–193. https://doi.org/10.2979/pme.2010.18.2.171

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft,* 9, 469–520. https://doi.org/10.1007/s11618-006-0165-2

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). Professional Competence of Teachers, Cognitively Activating Instruction, and the Development of Students' Mathematical Literacy (COACTIV): A research program. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Eds.), *Mathematics Teacher Education: Vol. 8. Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (pp. 1–21). Springer.

Bautista, A., Tan, C., Wong, J., & Conway, C. (2019). The role of classroom video in music teacher research: A review of the literature. *Music Education Research, 21*(4), 331–343. https://doi.org/10.1080/14613808.2019.1632278

Bell, C. A., Klieme, E., & Castellano, K. (2020). Social-emotional support. In Organisation for Economic Co-operation and Development [OECD], *Global Teaching InSights: A video study of teaching* (pp. 81–113). OECD Publishing. https://doi.org/10.1787/afaecf38-en

Bell, C. A., Klieme, E., & Praetorius, A.-K. (2020). S Conceptualising teaching quality into six domains for the Study. In Organisation for Economic Co-operation and Development [OECD], *Global Teaching InSights: Technical Report, Section I: Study background* (pp. 2–24). OECD Publishing. https://courtneybellphd.org/wp-content/uploads/2023/11/2020_GTI-TechReport-Chapter2.pdf

Björk, C., Hofmann, B., Knigge, J., & Puffer, G. (2022, November 25). *Towards modelling instructional quality for music classrooms: The relation between generic frameworks and subject-specific adaptations.* Poster presentation. 7th symposium of the research project "LeHet", University of Augsburg. Retrieved October, 04, 2024 from https://assets.uni-augsburg.de/media/filer_public/e9/31/e931704e-32a9-4ed8-911f-e06766fc588d/lehet-symposium_poster_bjork_etal.pdf

Björk, C., Ruthmann, S. A., Granfors, M., Högväg, J., & Andersson, S. (2021). The potential of a mixed-methods approach for research on learning to theorise music. *Music Education Research, 23*(3), 374–390. https://doi.org/10.1080/14613808.2020.1853085

Bowman, W. D. (2005). The limits and grounds of musical praxialism. In D. J. Elliott (Ed.), *Praxial music education: Reflections and dialogues* (pp. 52–78). Oxford University Press.

Brandstätter, U. (2008). *Grundfragen der Ästhetik. Bild – Musik – Sprache – Körper.* Böhlau.

Casey, C., & Moran, A. (1989). The computational metaphor and cognitive psychology. *The Irish Journal of Psychology, 10*(2), 143–161. https://doi.org/10.1080/03033910.1989.10557739

Charalambous, C.Y., & Litke, E. (2018). Studying instructional quality by using a content-specific lens: the case of the mathematical quality of instruction framework. *ZDM Mathematics Education, 50*(3), 445–460.

Concina, E. (2019). The role of metacognitive skills in music learning and performing: Theoretical features and educational implications. *Frontiers in Psychology, 10*(1583). https://doi.org/10.3389/fpsyg.2019.01583

Dewey, J. (1934). *Art as experience.* Minton, Balch & Co.; reprinted in *LW 10*.

Gabrielsson, A. (2011). *Strong experiences with music: Music is much more than just music*. (English ed.). Oxford University Press.

Gebauer, H. (2013). "Beschreibt doch mal die Form, die wir gerade gemacht haben." Kognitive Aktivierung im Musikunterricht ["Describe the musical form that we just played." Cognitive activation in the general music classroom]. In A. Lehmann-Wermser & M. Krause-Benz (Eds.), *Musiklehrer(-bildung) im Fokus musikpädagogischer Forschung* (pp. 61–79). Waxmann. https://doi.org/10.25656/01:9455

Godau, M. (2016). "Am besten ist, der Musiklehrer geht einen Kaffee trinken oder was weiß ich…" Zur Lehrer_innenrolle in selbständigen Lernprozessen im Musikunterricht ["The best thing would be if the music teacher left to have a coffee or whatever…". The teacher's role in self-directed learning processes in music lessons]. In J. Knigge & A. Niessen (Eds.), *Musikpädagogische Forschung: Vol. 37, Musikpädagogik und Erziehungswissenschaft* (pp. 156–170). Waxmann. https://doi.org/10.25656/01:14914

Greenberg, D. M., & Rentfrow, P. J. (2015). Rules of engagement: The structure of musical engagement and its personality underpinnings. In *Proceedings of the Ninth Triennial Conferences of the European Society for the Cognitive Sciences of Music*, Manchester, UK.

Haas, M., Nonte, S., Krieg, M., & Stubbe, T. C. (2019). Unterrichtsqualität in Musikklassen. Befunde aus der quasi-experimentellen Studie ProBiNi. In V. Weidner & C. Rolle (Eds.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung* (pp. 137–154). Waxmann. https://doi.org/10.25656/01:20701

Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education, 18*(1), 27–39.

Hasselhorn, J., & Knigge, J. (2021). Technology-based competency assessment in music education: the KOPRA-M and KoMus Tests. In A.

Lehmann-Wermser & A. Breiter, A. (Eds.), *Computer based assessment and feedback in music education* (pp. 49–74). Ifmpf. Retrieved October, 07, 2024, from https://www.researchgate.net/publication/353346184_Technology-Based_Competency_Assessment_in_Music_Education_the_KOPRA-M_and_KoMus_Tests

Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (5th ed.). Unterricht verbessern - Schule entwickeln. Klett/Kallmeyer.

Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: an integrative review and mapping of the literature. *Large-Scale Assessments in Education, 9*(17). https://doi.org/10.1186/s40536-021-00109-1

Herrmann, C., & Gerlach, E. (2020). Unterrichtsqualität im Fach Sport. Ein Überblicksbeitrag zum Forschungsstand in Theorie und Empirie. *Unterrichtswissenschaft, 48*(361–384).

Heß, F. (2018). Ästhetische Erfahrung. In M. Dartsch, J. Knigge, A. Niessen, F. Platz & C. Stöger (Eds.), *Handbuch Musikpädagogik. Grundlagen – Forschung – Diskurse* (pp. 181–187). Waxmann.

Hopfenbeck, T. N., & Lenkeit, J. (2018). *PIRLS for teachers: Making PIRLS results more useful for practitioners (Policy Brief No. 17).* IEA.

Juntunen, M.-L. (2015). National assessment meets teacher autonomy: National assessment of learning outcomes in music in Finnish basic education. *Music Education Research, 19*(1), 1–16. https://doi.org/10.1080/14613808.2015.1077799

Kertz-Welzel, A. (2015). Lessons from elsewhere? Comparative music education in times of globalization. *Philosophy of Music Education Review, 23*(1), 48–66.

Kleickmann, Th., Steffensky, M., & Praetorius, A.-K. (2020). Quality of teaching in science education: more than three basic dimensions? In A.-K. Praetorius, J. Grünkorn, J. & E. Klieme (Eds), Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen. *ZfP, 66. Beiheft* (pp. 37–55). Beltz.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. Aufgabenkultur und Unterrichtsgestaltung [Mathematics instruction at lower secondary level. Task culture and instructional design]. In Bundesministerium für Bildung und Forschung (BMBF) (Ed.), *TIMSS-Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43–57). BMBF.

Kranefeld, U. (2021). Der Diskurs um Unterrichtsqualität in der Musikdidaktik zwischen generischen und fachspezifischen Dimensionen. *Unterrichtswissenschaft, 49*(2), 221–233 . https://doi.org/10.1007/s42010-021-00113-y

Kranefeld, U., & Dücker, J. (2013). Unterrichtszeit nutzen. Inszenierungsmuster im instrumentalen Gruppenunterricht. In Koordinierungsstelle des BMBF-Forschungsschwerpunkts zu Jedem Kind ein Instrument (Ed.), *Empirische Bildungsforschung zu Jedem Kind ein Instrument. Ergebnisse des BMBF-Forschungsschwerpunkts zu den Aspekten Kooperation, Teilhabe und Teilnahme, Wirkung und Unterrichtsqualität* (pp. 72–75). Bielefeld.

Krause, M., & Oberhaus, L. (2012). *Musik und Gefühl. Interdisziplinäre Annäherungen in musikpädagogischer Perspektive.* Olms.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: the "cascade model". *ZDM Mathematics Education, 52*(2), 311–327. https://doi.org/10.1007/s11858-020-01151-z

Köller, O., Baumert, J., & Bos, W. (2001). TIMSS: Third International Mathematics and Science Study. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 269–284). Beltz.

Lietz, P., Cresswell, J. C., Rust, K. F., & Adams, R. J. (2017). Implementation of large-scale education assessments. In P. Lietz, J. C. Cresswell, K. F. Rust & R. J. Adams (Eds.), *Implementation of large-scale education assessments* (pp. 1–25). Wiley.

López-Íñiguez, G., & McPherson, G. E. (2020). Applying self-regulated learning and self-determination theory to optimize the performance of a concert cellist. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.00385

Lotz, M., Gabriel, K., & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung: Analysen zu deren gegenseitiger Validierung. *Zeitschrift für Pädagogik*, *59*(3), 357–380.

Mantie, R. (2016). Review of *A Brief Introduction to a Philosophy of Music and Music Education as Social Praxis*, by Thomas A. Regelski. *Philosophy of Music Education Review*, *24*(2), 213–219. https://doi.org/10.2979/philmusieducrevi.24.2.06

McPherson, G. E., & Renwick, J. M. (2001). A longitudinal study of self-regulation in children's musical practice. *Music Education Research, 3*, 169–186.

McPherson, G. E., Miksza, P., & Evans, P. (2017). Self-regulated learning in music practice and performance. In D. H. Schunk & J. A. Greene (Eds), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 181–193). Routledge.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center.

Möller, K. (2016). Bedingungen und Effekte qualitätsvollen Unterrichts - ein Beitrag aus fachdidaktischer Perspektive. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Eds.), *Bedingungen und Effekte guten Unterrichts* (pp. 43–64). Waxmann.

Niessen, A. (2010). Die Bedeutung von Verarbeitungstiefe im Musikunterricht – Dimensionen von Unterrichtsqualität einer Musikstunde [The importance of depth of processing in music education – dimensions of instructional quality in a music lesson]. In C. Wallbaum (Ed.), *Perspektiven der Musikdidaktik: Drei Schulstunden im Licht der Theorien* (pp. 63–82). Olms.

Odena, O. (2018). Creativity in the secondary music classroom. In G. McPherson & G. F. Welch (Eds.), *The Oxford handbook of music education, Volume 2. Music learning and teaching in infancy, childhood, and adolescence* (pp. 295–312). Oxford University Press.

Organisation for Economic Co-operation and Development [OECD] (2017). *TALIS 2018 Video study and global video library on teaching practices.*

Organisation for Economic Co-operation and Development [OECD] (2018). *Global Teaching InSights, Technical Report, Annex A. TALIS video observation codes: Holistic domain ratings and components, Version 5.0.*

Organisation for Economic Co-operation and Development [OECD] (2019). *TALIS 2018 Technical report*. Teaching and Learning International Survey.

Organisation for Economic Cooperation and Development [OECD] (2020). *Global Teaching InSights: A video study of teaching*. OECD Publishing. https://doi.org/10.1787/20d6f36b-en

O'Neill, S. (2012). Becoming a music learner: Toward a theory of transformative music engagement. In G. McPherson & G. F. Welch (Eds.), *The Oxford handbook of music education, Volume 1* (pp. 163–186). Oxford University Press.

Opfer, V. D. (2020). The rationale of the Study. In OECD, *Global Teaching InSights: A Video Study of Teaching* (pp. 17–32). OECD Publishing. https://doi.org/10.1787/9d9b4005-en

Opfer, V. D., Bell, C. A., Klieme, E., McCaffrey, D. F., Schweig, J., & Stecher, B. M. (2020). Understanding and measuring mathematics teaching practice. In OECD, *Global Teaching InSights: A video study of teaching* (pp. 33–47). OECD Publishing. https://doi.org/10.1787/98e0105a-en

Orzolek, D. C. (2019). Music teacher evaluation: An overview. In T. S. Brophy (Ed.), *The Oxford Handbook of assessment policy and practice in music education* (pp. 795–923). Oxford University Press.

Praetorius, A.-K., Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM Mathematics Education, 50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B., & Nehring, A. (Eds.).(2020a). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft, 48*(3), 409–446. https://doi.org/10.1007/s42010-020-00082-8

Praetorius, A.-K, Klieme, E., Kleickmann, Th., Brunner, E., Lindmeier, A., Taut, S., & Charalambous, C. (2020b). Towards developing a theory of generic teaching quality: Origin, current status, and necessary next steps regarding the three basic dimensions mode. In A.-K. Praetorius, J. Grünkorn & E. Klieme (Eds), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen. ZfP, 66. Beiheft*, 15–36. Beltz.

Ray, J. (2004). *Musikaliska möten man minns: om musikundervisningen i årskurserna sju till nio som en arena för starka musikupplevelser*. [Memorable musical encounters. About music education for teenage pupils in compulsory school as an arena for strong experiences related to music.] [Doctoral dissertation, Åbo Akademi University]. Retreived October, 07, 2024, from https://www.doria.fi/bitstream/handle/10024/4139/TMP.objres.43.pdf?sequence=2

Rolle, C. (1999). *Musikalisch-ästhetische Bildung: Über die Bedeutung ästhetischer Erfahrung für musikalische Bildungsprozesse* [Musical-aesthetic education. On the importance of aesthetic experience for musical education processes]. Bosse.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Seel, M. (1985). *Die Kunst der Entzweiung. Zum Begriff der ästhetischen Rationalität.* Suhrkamp.

Seel, M. (2005). *Aesthetics of appearing (Cultural memory in the present)*. Stanford University Press. (First published in 2000: Ästhetik des Erscheinens. Hanser.)

Steinbach, A. (2018). Musikalische Sozialisation und soziale Dimensionen des Musiklernens. In M. Dartsch, J. Knigge, A. Niessen, F. Platz, & C. Stöger (Eds.), *Handbuch Musikpädagogik: Grundlagen – Forschung – Diskurse* (pp. 228–235). Waxmann.

TVS International Consortium (2019). *The TALIS video study observation system.*

Uljens, M., & Ylimaki, R. M. (2017). Non-affirmative theory of education as a foundation for curriculum studies, Didaktik and educational leadership. In M. Uljens & R. M. Ylimäki (Eds), *Bridging educational leadership, curriculum theory and Didaktik* (pp. 3–145). Springer. https://doi.org/10.1007/978-3-319-58650-2_1

van Tartwijk, J., & Hammerness, K. (2011). The neglected role of classroom management in teacher education. *Teaching Education, 22*(2), 109–112. https://dx.doi.org/10.1080/10476210.2011.567836

Vogt, J. (2003). Philosophy–Music Education–Curriculum: Some casual remarks on some basic concepts. *Action, Criticism & Theory for Music Education, 2*(1), 1–25. https://Act.maydaygroup.org/articles/Vogt2_1.pdf

Vogt, J. (2017). Music education and ideology critique – a new approach. *Zeitschrift für neue Musikpädagogik,* 35–51. Retrieved October, 07, 2024 from https://www.zfkm.org/17-vogt1.pdf

Wagemaker, H. (2014). International large-scale assessments: From research to policy. In L. Rutkowski, M. von Davier & D. Rutkowski (Eds.), *Statistics in the social and behavioral sciences series. Handbook of international large-scale assessment. Background, technical issues, and methods of data analysis* (pp. 11–36). CRC Press.

Wallbaum, C. (2018a). *RED – A supposedly universal quality as the core of music education*. In C. Wallbaum (Ed.), *Comparing international music lessons on video* (pp. 123–146). Georg Olms.

Wallbaum, C. (Ed.). (2018b). *Comparing international music lessons on video*. Georg Olms.

Westerlund, H. (2002). *Bridging experience, action, and culture in music education*. [Doctoral dissertation, Sibelius Academy]. Retrieved October, 07, 2024, from https://urn.fi/URN:ISBN:952-9658-98-2

Wiggins, J. (2015). *Teaching for musical understanding* (3rd ed.). Oxford University Press.

Wiggins, J., & Espeland, M. I. (2018). Creating in music learning contexts. In G. McPherson & G. F. Welch (Eds.), *The Oxford handbook of music education: Volume 2. Music learning and teaching in infancy, childhood, and adolescence* (pp. 121–140). Oxford University Press.

Zandén, O. (2018). Enacted possibilities for learning in goals- and results-based music teaching. In C. Wallbaum (Ed.), *Comparing international*

music lessons on video (pp. 255–274). Georg Olms.

Zhu, Y., & Kaiser, G. (2022). Impacts of classroom teaching practices on students' mathematics learning interest, mathematics self-efficacy and mathematics test achievements: a secondary analysis of Shanghai data from the interna tional video study Global Teaching InSights. *ZDM Mathematics Education, 54,* 581–593. https://doi.org/10.1007/s11858-022-01343-9

Funding

This work was supported by the Swedish Cultural Foundation in Finland [grant number 158935].

Authors

Cecilia Björk
Universität für Musik und darstellende Kunst Wien
Åbo Akademi University
bjoerk@mdw.ac.at

Bernhard Hofmann
Universität Augsburg
bh@mupaed.de

Jens Knigge
Nord University, Campus Levanger
jens.knigge@nord.no

Gabriele Puffer
Hochschule für Musik und Theater München
gabriele.puffer@hmtm

Björk et al., 2024
CC BY-NC 4.0
b-em (16), S. 1-17

ISSN: 2190-317
DOI: https://doi.org/10.62563/bem.v16i2.246