Andreas Lehmann-Wermser & Andreas Breiter (Hrsg.)

Computer Based Assessment and Feedback in Music Education

Dokumentation des Abschlusssymposiums im Forschungsprojekt *PosyMus*

Institut für musikpädagogische Forschung
Hochschule für Musik, Theater und Medien Hannover

Forschungsbericht Nr. 29

Veröffentlichungen des Instituts für musikpädagogische Forschung der Hochschule für Musik, Theater und Medien Hannover

Forschungsbericht
herausgegeben von Andreas Lehmann-Wermser

der Hochschule für Musik, Theater und Medien Hannover
http://www.ifmpf.hmtm-hannover.de/
e-mail: ifmpf@hmtm-hannover.de

Bestellungen über die Homepage des Institutes.

Redaktion: Kaja Hahnheiser

Umschlaggestaltung: Frank Heymann

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

Hannover: ifmpf 2020

ISSN 1430-8088
ISBN 978-3-931852-51-1

Institut für musikpädagogische Forschung
Hochschule für Musik, Theater und Medien Hannover

Andreas Lehmann-Wermser & Andreas Breiter (Eds.)

Computer Based Assessment and Feedback in Music Education

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Forschungsbericht Nr. 29

Hannover 2020

Inhalt

Andreas Breiter[1] & Andreas Lehmann-Wermser[2]

1. Potentials of Systematic Feedback in Music

Assessment in its many forms has become one of the key terms when one talks about development (and especially improvement) in the educational sector. In 1997 the decision of the German Standing Commission of the State's Ministers for Education (KMK) to participate in large scale assessments (LSA) like TIMSS and PISA and the following publication of devastating results have triggered a structural change of unprecedented depth: financial commitments by the Federal government (that needed constitutional changes), the expansion of all-day-schools (that needed extended buildings) or new forms of cooperation in so-called "multi-professional teams" (Speck et.al., 2011) that also included the aesthetic subjects (Lehmann-Wermser et al., 2010) are just a few of the results that are direct or indirect consequences of the large scale assessments. Along came an equally dramatic shift in research as large research programs were set up to accompany all measures (Fischer & Holtappels, 2011) in order to provide a solid knowledge basis for future changes.

While all these changes referred to the macro or meso level of educational governance, i. e. to changes in the administration and organization of learning processes on the policy level, attention was also given to the micro level, i. e. the learning itself. What is happening in class? How can teaching (and subsequently learning) be changed to be more effective, more just, more profound? A change of paradigms took place leading to a different understanding of the role of empirical research. Student assessments were now considered an important element of school improvement, classroom management and development and learning support. All measures are part of a larger

[1] Institute for Information Management Bremen GmbH (ifib)

[2] Hannover University of Music Drama and Media

context in which assessment of any kind of performance is not only one indicator but also a tool for future learning processes for the use of both students and teachers. This important shift has been coined as the change from assessment *of* learning (as in the case of LSA) to assessment *for* learning which is the focus of the project to be documented here (Fautley, 2010). Feedback is the cornerstone within this concept may it be data driven or less structured, it has been listed as a core feature of good teaching in general (Kiel 2018), which is supported by Hattie's meta study (Hattie & Timperley, 2007). The KMK describes a "feedback culture as the interface between data monitoring and data use" (KMK 2010, p. 17). But feedback relies on the existence of valid and reliable data.

How to get data? What to do with data?

Those who proclaim data driven school development (Mandinach, 2012) idealistically assume a chain that starts with reasonable data collection (on various levels), an analysis of these collected data followed by an interpretation of the data (taking into respect interests and perspectives of the acting people on various levels), and a feedback to the actors (adjusted to their needs and capabilities) in order to take measures to develop or improve.

Obviously, the intended mechanism has not always worked out that way. One of the crucial points to take into consideration is how the collected data must be read, interpreted and translated into individual support measures by teachers. The interface between test results and targeted teacher action is of particular interest in the subject of music education.

Unlike other main subjects ("core subjects"), music in Germany as a small, aesthetic subject has not participated in large scale assessments[3]. Hence, after

[3] Here, „large scale" is understood as nationwide or even international assessments oftentimes constructed, conducted and evaluated by international teams. However, there are a few exceptions of music specific LSA on state level, mostly in the USA, like the South Carolina Arts Assessment Program (SCASS 1999) or in Florida (Brophy, T. S. The Florida Music Assessment Project: An Association Driven Model of Large Scale Assessment

the turn of the century there was no effort in creating valid and reliable test items by the big test management consortia that could stand up to the internationally established psychometric standards. Furthermore, at least in Continental Europe music education was regarded as a creative and design-oriented subject in which standardized tests would not work for reasons following the didactical structure of the discourse. Fortunately, by now there is some music specific research on test construction and validation, some of which is feeding this documentation of the symposium. Assessment instruments are available for several sub-areas of behaviour relevant to teaching and learning music and are presented in this volume. But these tests were not developed for the classroom and everyday use by teachers and learners. This, however, is crucial for the development of the subject (Breiter & Light, 2006).

Independently from music as a subject, quite a body of research has been devoted to find out how make use of the test results. In most of the existing research, the focus was on general aspects of feedback systems, like timeliness, quality, explanation etc. One focus was also on test data literacy of teachers (Gummer & Mandinach, 2015) how can they interpret the data to improve teaching and learning. As there are no subject specific findings on how to deal with feedback from the test results, at least two research question arise in the case of music: firstly, how to create computer-based test instruments for music education which are accessible in every classroom and secondly, how to build a user-centered web-based interface to visualize test data and to help teachers (and learners) to understand the data (Visscher & Coe, 2003; Breiter & Stauke, 2007).

Development. Retrieved from https://flmusiced.org/dnn/ Portals/0/ MusicAssesProject/PDF/Brophy_assessment.pdf

The [PosyMus] Project

In a 3-year research project financed by the Federal Ministry for Research and Education (BMBF[4]), researchers from Music Education and Informatics have collaborated to create a usable web-based tool for testing students in music and to support teachers with an information system for performance feedback[5]. The project used test items that have been validated in research projects. The KoMus project (Niessen et al., 2008, Jordan et al., 2012) had modelled competencies in the field of listening to and understanding of music. Likewise, Kopra-M (Hasselhorn & Lehmann, 2014) covered the field of melodic and rhythmic competencies both on instruments and while singing.

A number of the items from these projects were translated into a uniform format compatible with the conditions at secondary schools. Computer-based tests were done on tablet computers. For this purpose, 10 secondary schools in Germany were selected that could provide the web capacity needed. In addition, it was expected that teachers and students should be familiar with digital tools and show basic competencies. In order to control for context-specific criteria, attitudes towards and competences in dealing with digital media and with feedback from performance tests were surveyed among the participating teachers. Building on this, an electronic feedback system (eRMS) was developed and designed together with the teachers to meet the multimedia requirements of music. In a mixed methods approach, the technical and personal conditions were evaluated, focusing in particular on the use of the eRMS to support the individual development of students. At the same time, the possibilities of teacher training in the subject of music through the use of digital media were explored. The aim was to further develop forms

[4] Registration number 01LSA1506A

[5] Involved in the project were Franziska Marx, Michaela Meyer, Tammo Gehrig and Sebastian Ritter on the side of the ifib.

of multimedia testing of music skills that are suitable for large-scale assessments and can be transferred to other subjects if necessary.

About this book

This book documents some of the presentation given at a symposium held at the University of Music Drama and Media Hannover on September 15, 2017. It gathered researchers from several academic disciplines (see contributor's list in the appendix; see also the complete program in the appendix).

The book opens with two chapters documenting the design and results of the project: **Julia Finken** is focussing the information science aspects while **Andreas Lehmann-Wermser** and **Benjamin Weyel** show the possible use of computer based feedback in the music classroom.

Feedback needs a solid base in data collection. [PosyMus] could rely on tests developed by **Jens Knigge** and **Johannes Hasselhorn**. They both started working on the modelling of competencies in music in research projects financed by the German Research Foundation (DFG). The items they generated in their tests laid the ground for [PosyMus] as they provided psychometrically sound problems that could be adapted for our project. They sketch the background for competence models in music education and for computer-based testing in general.

Marina Gall brings in the international perspective as feedback and assessment can be seen in a wider context. The English school system lays out quite a different context for any approaches of feedback, so the exchange can be quite stimulating.

References

Breiter, A. & Light, D. (2006). Data for school improvement: factors for designing effective information systems to support decision-making in schools. *Educational Technology & Society*, *9*(3), 206-217.

Breiter, A. & Stauke, E. (2007). Anforderungen an elektronische Rückmeldesysteme aus Nutzersicht. *Empirische Pädagogik*, *21*(4), 383-400.

Brophy, T. S. (2008). The Florida Music Assessment Project: An Association Driven Model of Large Scale Assessment Development. In T.S. Brophy & K. Albert. (Eds.), *Assessment in Music Education: Integrating Curriculum, Theory, and Practice. Proceedings of the 2007 Florida Symposium on Assessment in Music Education* (pp. 139-151), Chicago: GIA.

Fautley, M. (2010). *Assessment in Music Education*. Oxford, New York: Oxford Univ. Press.

Fischer, N. & Holtappels, H. G. (Eds.) (2011). *Ganztagsschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagsschulen (StEG). Studien zur ganztägigen Bildung*. Weinheim: Juventa.

Gummer, E. S. & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, *117*(4), 1-22.

Hasselhorn, J. & Lehmann, A. C. (2014). Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz. In B. Clausen (Ed.): *Teilhabe und Gerechtigkeit* (pp. 79-95) Münster: Waxmann.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112.

Jordan, A.-K., Knigge, J, Lehmann, A. C., Niessen, A. & Lehmann-Wermser, A. (2012). Entwicklung und Validierung eines

Kompetenzmodells im Fach Musik – Wahrnehmen und Kontextualisieren von Musik. *Zeitschrift für Pädagogik*, *58*(4), 500-521.

KMK – Kultusministerkonferenz (2010). Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung. Retrieved from: http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2009/2009_12_ 10-Konzeption-Bildungsstandards-Bs.pdf [29.01.2016].

Lehmann-Wermser, A., Naacke, S., Nonte, S. & Ritter, B. (Eds.) (2010). *Musisch-kulturelle Bildung an Ganztagsschulen: Empirische Befunde, Chancen und Perspektiven. Studien zur ganztägigen Bildung*. Weinheim [u.a.]: Juventa-Verl.

Mandinach, E. B. (2012). A perfect time for data use. Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71-85.

Niessen, A., Lehmann-Wermser, A., Knigge, J. & Lehmann, A. C. (2008): *Entwurf eines Kompetenzmodells "Musik wahrnehmen und kontextualisieren" (special volume), 3-33*. Retrieved from: http://www.zfkm.org/sonder08-niessenetal.pdf [30.05.2008].

Speck, K., Olk, T. & Stimpel, T. (2011). Auf dem Weg zu multiprofessionellen Organisationen? Die Kooperation von Sozialpädagogen und Lehrkräften im schulischen Ganztag. Empirische Befunde aus der Ganztagsforschung und dem Forschungsprojekt "Professionelle Kooperation von unterschiedlichen Berufskulturen und an Ganztagsschulen. *Zeitschrift für Pädagogik*, *57*, 184-201, Weinheim: Beltz.

Visscher, A. J. & Coe, R. (2003). School performance feedback systems. Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement*, *14*(3), 321–349.

Julia Finken[1]

2. [PosyMus] - the Technical Perspective

One of our goals in [PosyMus] was to develop a testing and feedback system that could be used on tablets for flexible use in classrooms which are intuitive and easy-to-use. On the technical side, this goal could be split into two sub-goals: the development of a computer-based test system (CBA) for students and the development of a feedback system for teachers. The test system was supposed to be used in a tablet environment; meanwhile, the feedback system was optimized for desktop use. In the following article, we describe the two systems, including their development, testing and the feedback from the users.

Test System

For building the base of our test system we selected 27 items out of a pool of test items that were developed and validated in two former research projects: KOMUS (Jordan et al., 2010) and KOPRA-M (Hasselhorn, 2015; Hasselhorn & Lehmann, 2014). These former items were optimized for desktop use and partially based on Flash. As we wanted to use tablets in our project, these conditions did not fulfill our requirements so we were not able to use the original items but had to re-program and re-design them, inorder to make them more suitable for our application. This also implicated to make a restriction regarding the type of items that we wanted to use. The already developed and validated items contained the dimensions of contextual knowledge, melody, rhythm and singing. We were not able to include and analyze the singing test items because of their structure and their requirements for realizing them in class, so we decided to skip these and only use items of the other three dimensions.

[1] Institute for Information Management Bremen GmbH (ifib)

We had to consider different aspects of being able to decide which framework we wanted to use for our test system (Finken et al., 2016). For the implementation of the items, we needed to include the possibility to listen to audio files and to record keystrokes on a touchscreen piano made by the users. With two years for the whole project, there was not that much time for developing a completely new framework. Therefore, we decided that it would make more sense to use an existing framework which offers the opportunity to deal with audio files and input. Another reason for that was the possible acceptance of the context of large-scale assessments. An already practically proven solution might have better chances to be used than a newly developed framework.

For choosing a suitable framework we had to consider some requirements (Finken et al., 2016): the system must be usable on mobile devices, should be platform independent, open source, accessible via the internet, should save the privacy of the users and be suitable for doing large-scale assessments. If possible, it should support accessibility. Also implementing and changing the items should be easy. And finally, it needed to support multimedia items. Therefore we compared possible solutions such as Tao, Illias or Stud.IP.

After comparing these frameworks, we decided to use Tao which was also used for the PISA (Programme for International Student Assessment) tests. Tao is platform independent and usable via the internet. We optimized our system for Chrome and Safari because those are common browsers on tablets. Tao is account based, so it could be used for groups. For fulfilling the data privacy condition and not being able to correlate students and accounts we used anonymized accounts. So, only the teacher of a class is able to track the results in the feedback system to the students. Tao offers two options for building items: QTI and OWI. Because of the use of multimedia items and therefore the need for having various options regarding audio playing and

recording we had to use OWI items. That meant we had to build all items on our own with HTML, CSS and Javascript.

One of the main developments for the items was the so-called "Color Piano" (see figure 2, p. 25). This is a further development of the "Colored Music Grid App" of Hasselhorn (2015). It is built with HTML, CSS and Javascript. It uses JS-libraries for being able to play sounds on mobile devices and record the keystrokes on the color piano made by the students. We tried to create the same conditions for students which have knowledge about playing musical instruments and playing on sight and those without such knowledge. So, the sounds in the Color Piano are represented by colours. There are also notes and bars but only for showing the length of a sound and for easing the orientation. For helping the students not to lose track of the actual note we implemented a slider (see the black "[" in the picture ...) to mark the current position in the song. We also added a counting in and a gong for marking the end of the practise phase. During playing there is a reference melody for making the orientation easier. We tried to consider accessibility issues and therefore used a responsive design and tried to make the colours accessible.

In the following, there are examples of the three types of items: contextual, melody, rhythm (see figure 2&3, p. 25).

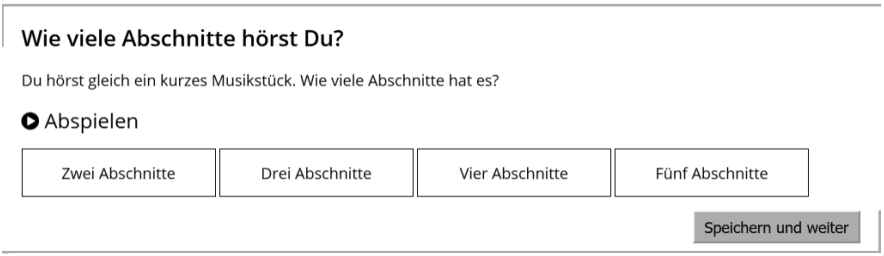


Fig. 1: contextual item

Test procedure in schools

Five schools participated in our project. These provided us with the possibility to do our test with ten classes and about 220 students from the grades 6, 7 and 9. The tests took place during a normal lesson. In the beginning, the project team gave a small introduction to the project, the meaning of the test and the test system. It was emphasised, that the students would not get a grade for the test to remove any pressure. After the introduction, the students started with a small test, with just two items, in order to familiarise themselves with the use of the Color Piano. During the development phase the project team noticed that using the Color Piano for the very first time is not easy, so the students got this practising period to minimize the disadvantages of being a first-timer. After the practising phase, the students were asked to start with the actual test. They needed about 20-45 minutes for finishing it. If possible, the project team did a review with the students and the teacher after the test. They asked e.g. about what the students liked and did not like about the test or if the items were difficult or easy to handle. These reviews made some difficulties with the test items visible, thus allowing the project team to change the mentioned items for the next test in the next school, e.g. lowering the volume of the background music during the rhythm items or to add some additional items for fast students so that they have something to do while waiting for the other students to finish the test (Finken et al., 2017).

Test setup

German schools are somewhat dragging behind in their use of digital media; hardware and software are below standards in the Western countries. We provided two technical solutions for doing the test in the schools: a virtual server, accessible via the internet, and a local server, accessible via a WiFi router. After doing a technical interview with the participating schools considering the technical infrastructure in the schools, we decided to use the local server solution (Finken et al., 2017). In some schools the internet and

Wifi accessibility was too insufficient (too slow, not covering the whole area) so accessing our test in real-time without delay would not have been possible by using the first solution. Another benefit was the "missing opportunity" to use the internet for doing other things as looking for answers or chatting with other students. So there was less distraction for everyone.

For the setup in the schools we used al local server and a WiFi-Router to build up a local network (see figure 4). The students connected their tablets to this local server. They used headphones to be less distracted by noise.

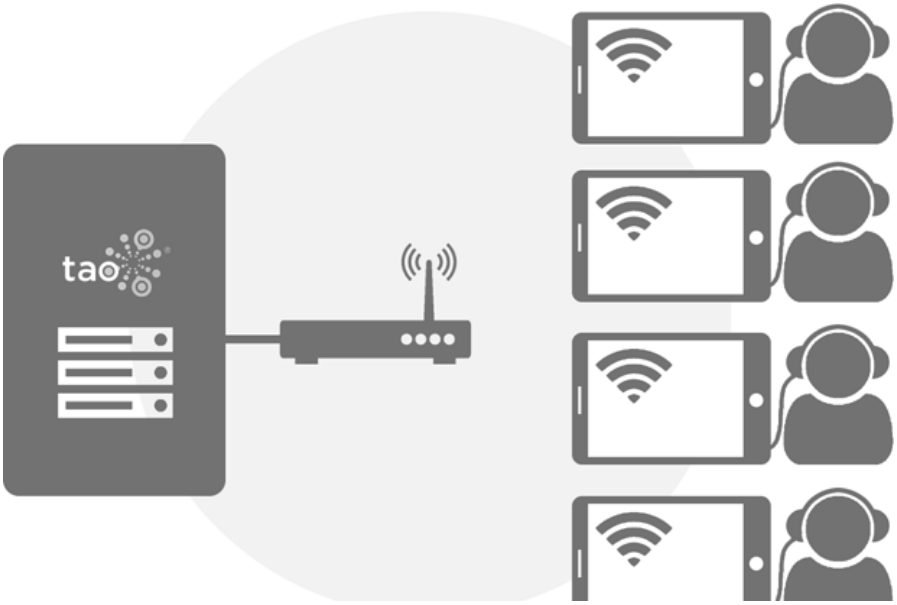


Fig. 4: Technical setup in schools

Results of the tests

The results of the tests were similar in all the participating schools (Finken et al., 2017). Some of them could even be used to optimize the items for the

next test in the next school. The following list shows some feedback we received from students.

- The background music in the rhythm items was too loud, the students were not able to hear their own version of the rhythm.
- The speed of the melody items was too fast.
- 10 seconds for practising is not enough.
- Some students were confused by the changing order of the colours.
- The length of the test was described as appropriate.
- Tasks were estimated as "achievable".
- The students would have liked to get direct feedback after finishing an item.
- They had fun using the test system.

Altogether administering our computer based test in schools turned out to be manageable and convenient for everyone.

Feedback system

The idea behind the feedback system was to provide detailed feedback about the performance of each student back to the teacher to help them organize and optimize their lessons.

Just like the test system, the feedback system has been implemented with HTML, CSS and Javascript. For visualising charts and diagrams we decided to use „Highcharts“, a Javascript library for interactive visualisation. The feedback system, too, is web based and platform independent. In contrast to the test system, it is optimized for desktop use because of the possibilities that a bigger display offers for effective visualisations.

To visualise the results of the tests, we need to export them out of Tao as a CSV-file and load them into our feedback system manually. The data are held locally in the browser. Because of that, the teachers need to import the results again if they close the browser or want to compare the results of e.g.

two years. This has been done because of privacy reasons: the collected data are quite sensitive, although they are anonymized.

In the feedback system, the results are analyzed and displayed automatically. The test result data of the students is transferred into visualisations. Those visualisations are either interactive or non-interactive. It is important to say, that the feedback system only visualises the results, there is no grading included. Only the teachers are supposed to do that.

Development process

Our development process was following the User-Centered Design process (Nielsen, 1993; Abras et al., 2004; Gould & Lewis, 1985, Yau, 2013). We started to discuss the first designs on the base of mockups and paper prototyping (Sarodnick & Brau, 2011; Zeising & Katterfeldt, 2013) with the project team and also with teachers. We also used some low-level HTML prototypes. The use of these two methods was not as suitable as we thought. It was difficult for the teachers, who were not used to these design approaches, to think about the usefulness of a visualisation without deeper knowledge of the test items. Also, it was hard for them to think about the interactivity of a visualisation as long as it was not interactive. So we changed the setting over the course of time. On the one hand, we gave the teachers a detailed introduction to the test system. On the other hand, we did usability tests with prototypes of a higher fidelity and gave them a set of tasks for working with the system. During these usability tests, we used the question asking protocol method (Barnum, 2011; Nielsen, 2010) along with screen capturing and voice recording. Thus we were able to analyze the problems later on a base of different sources and were able to integrate the thoughts and opinions of the teachers into the development of the system.

Design and structure of the feedback system

The design of the feedback system was based on two major decisions: the visualisation should look as similar to the test system as possible and the visualisations of the pracice items should be interactive. The first decision was made to make it easier to recognize the test items and interpret the results. We also decided to highlight the correct, wrong, yet plausible sounding, and the students answer directly in the visualisation of the item. The advantage of interactive items is the possibility to listen to the results of the students as often as the teachers want to. It also is possible to compare the results of the students with the original items via listening.

Based on these decisions we developed four different views for structuring the system:

1. an overview of all items for the class
2. an overview of all items for one student compared to the class
3. an overview of one item for the class
4. a single view of one item for one student.

The first view provides the teacher with a first idea of how the students performed in general. It lists all the items of the test and the summed up results of the class in the form of a bar chart. The teacher can compare the results of a single student against the overall performance which are shown next to each other, by selecting the second tab (view 2) (see figure 5, p. 26).

View 3 and 4 differ according to the type of item, contextual or practical. For the contextual items we used classical visualization types, like bar and pie charts, for the class view of one item. The correct answer is shown in the sidebar (see figure 6, p. 27).

The visualisation of the view for a single student is based on the original design of the item. The answer of the student is marked in the item and also explicitly listed in the sidebar. This is also done for the correct answer and the most plausible sounding wrong answer. This means the answer is wrong

but very close to being correct. The answer of the student is colored in blue, the correct answer is outlined in green and the most plausible wrong answer is outlined in orange (see figure 7, p. 27).

The single view of one item and one student shows the original note lines along with the lines that the student played. They are paired up one underneath the other so that it is possible to see which note a student played at every moment in the item. While playing the visualisation a slider is marking

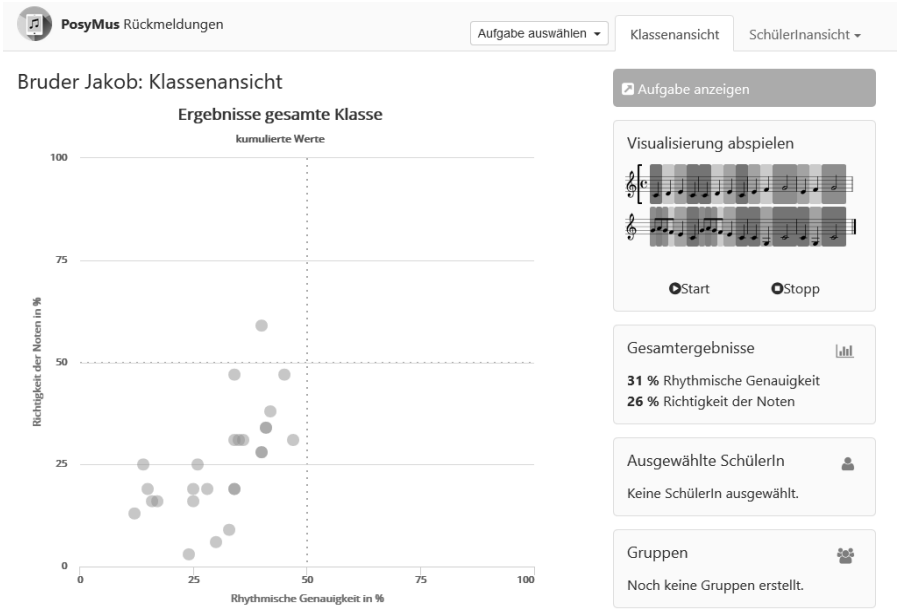


Fig. 8: Contextual item - Where are the mistakes?

the current note, so one does not loose track, in the same way as it is done in the original test item. The visualisation can be paused to have a closer look at one specific moment. We also listed the average correctness of rhythm and

notes in the sidebar for the class and the student to make it easier to compare a student to the class.

For the class view of one practical item we decided to use a visualization that shows the development of the class during the item. So we constructed a coordinate system build on the rhythmic and the tonal accuracy. Every student is visualized as a dot in the coordinate system. In the sidebar we integrated a player showing and playing the item. This player can be started and paused by the teacher at any time. During playing the music the dots in the

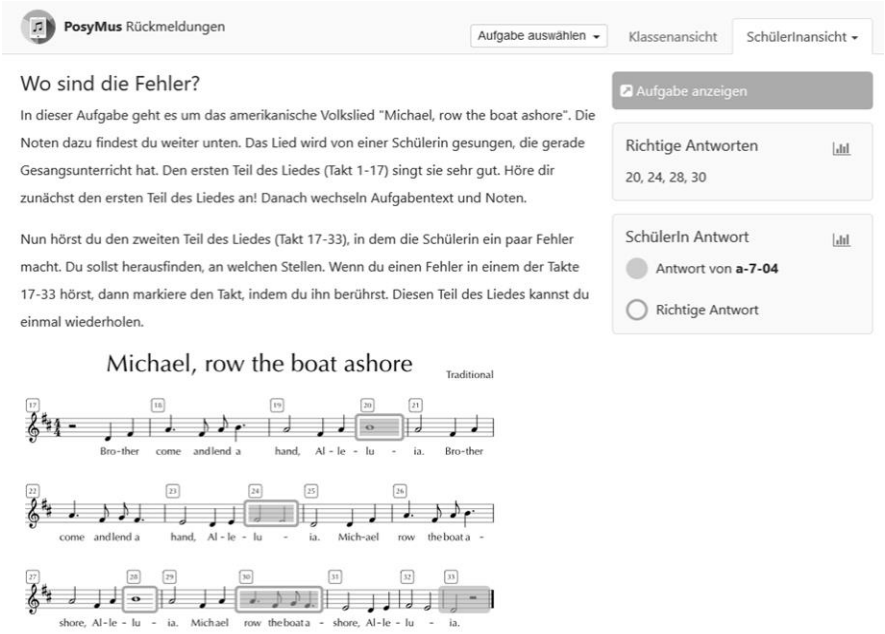


Fig. 10: Class view for practical items

coordinate system move according to the cummulated values for both dimensions for the students at that note. Therefore, it is possible to see how the students perform throughout the item, compared to the other students of a

class. It is also possible to see if there are notes in the item that are more difficult than other notes. The overall accuracy of rhythm and notes for the class is listed in the sidebar (see figure 9, p. 28). Another feature is the possibility to build groups of students and to see how these perform. They are highlighted in the coordinate system via a chosen colour. So it is possible to compare e.g. students that play an instrument with students that do not play an instrument.

Results of the usability tests

The usability tests with the teachers gave us some interesting points to consider for the development and also an estimation of the usefulness of our system:

- The information about the average class and student performance for an item in the side boxes are helpful for comparing.
- Working with the interactive visualisations for the single student views is easy.
- The class overview for the music practice items are more complicated to understand in the beginning, but rated as easy and helpful afterwards.
- Most of the chosen visualisations are appropriate to indicate the performance of students using the test system.
- The used classic charts (bar, pie) are easy to interpret.

As a conclusion we can say that most of the chosen visualizations are appropriate to show the performance of the entire class as well as a single student. The teachers found the idea of getting help in rating the students more objectivly very interesting. They were able to recognize problems their students had with certain tasks and so they could take the outcome of the tests to deepen that topic in their lessons.

But: For being able to grade students is it important to know them and what they are good in and where they have learning difficulties. So our system is not rating the students it just visualizes the results.

References

Abras, C., Maloney-Krichmar, & D., Preece, J. (2004). User-centered design. In W. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction*, *37*(4), (pp. 445-456). London: Thousand Oaks, Sage Publications.

Barnum, C. M. (2011). *Usability testing essentials: ready, set – test*. Burlington, MA: Morgan Kaufmann Publishers.

Finken, J., Marx, F. & Breiter, A. (2016). Computerbasiertes Testen zur Messung von Musikkompetenzen [Computer based testing of musical comtetences]. In U. Lucke, A. Schwill, & R. Zender (Eds.), *Proceedings of DeLFI 2016 – Die 14. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e. V.* (pp. 327-329). Bonn: Gesellschaft für Informatik.

Finken, J., Marx, F., Meyer, M., Krieter, P., & Breiter, A. (2017). Entwicklung und Durchführung computerbasierter Tests zur Messung von Musikkompetenzen [Development and administration of computer based tests of musical competences]. In C. Igel, C. Ullrich, & M. Wessner (Eds.). *Bildungsräume 2017. Proceedings of DeLFI 2017 - Die 15. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e. V.* (pp. 63-74). Bonn: Gesellschaft für Informatik.

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.

Hasselhorn, J. (2015). *Messbarkeit musikpraktischer Kompetenzen von Schülerinnen und Schülern: Entwicklung und empirische Validierung eines Kompetenzmodells* [Measurability of students‘ practical competences in music]. Münster: Waxmann.

Hasselhorn, J., & Lehmann, A. (2014). Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz (KOPRA-M) [Development of a validated model od practical competences in music].

In B. Clausen (Ed.), *Teilhabe und Gerechtigkeit* (pp. 77–94). Münster: Waxmann.

Jordan, A.-K., Knigge, J., & Lehmann-Wermser, A. (2010). Projekt KoMus: Entwicklung von Kompetenzmodellen in einem ästhetischen Fach [Project KoMus: Development of competence models in an esthetic subject]. In A. Gehrmann, U. Hericks, & M. Luders (Eds.), *Bildungsstandards und Kompetenzmodelle - Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (pp. 209–222). Bad Heilbrunn: Klinkhardt.

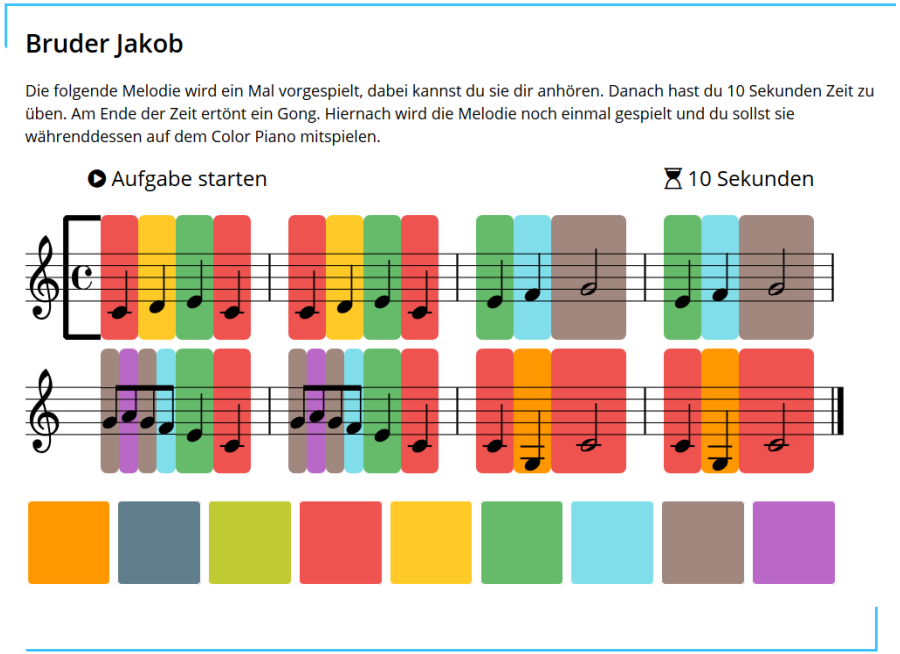
Nielsen, J. (1993). *Usability Engineering*. San Francisco; Kaufmann.

Nielsen, J. (2010). *Usability Engineering*. Nachdr. Amsterdam: Kaufmann.

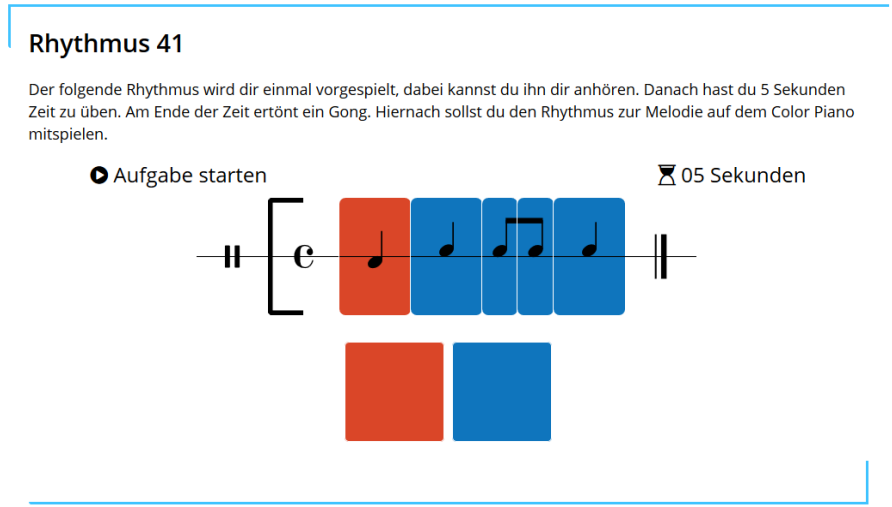
Sarodnick, F., & Brau, H. (2011*). Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung* [Methods of usability evaluation: scientific basics and practical applications]. Bern: Huber.

Yau, N. (2013). *Data points: Visualization that means something*. John Wiley & Sons.

Zeising, A., & Katterfeldt, E.S. (2013). *Where Is the "Like" Button? Going Beyond Usability when Designing for and with Teens. Workshop "Methods of Working with Teenagers in Interaction Design"* at CHI 2013. Paris. Retrieved from: www.chici.org/teens2013/papers/chi2013-w22-paper13[12.10.2018].

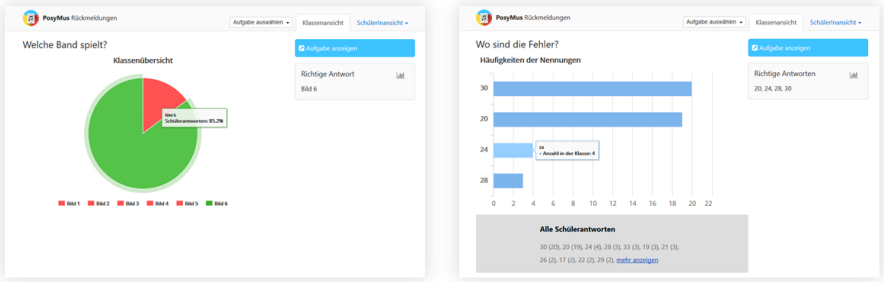


Finken J.: Fig. 2: melody item

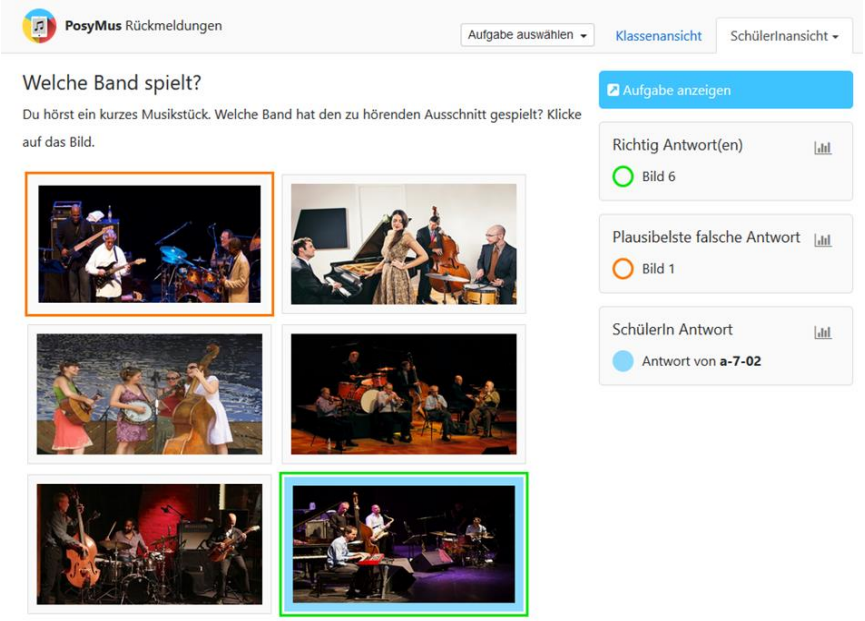


Finken J.: Fig. 3: rhythm item

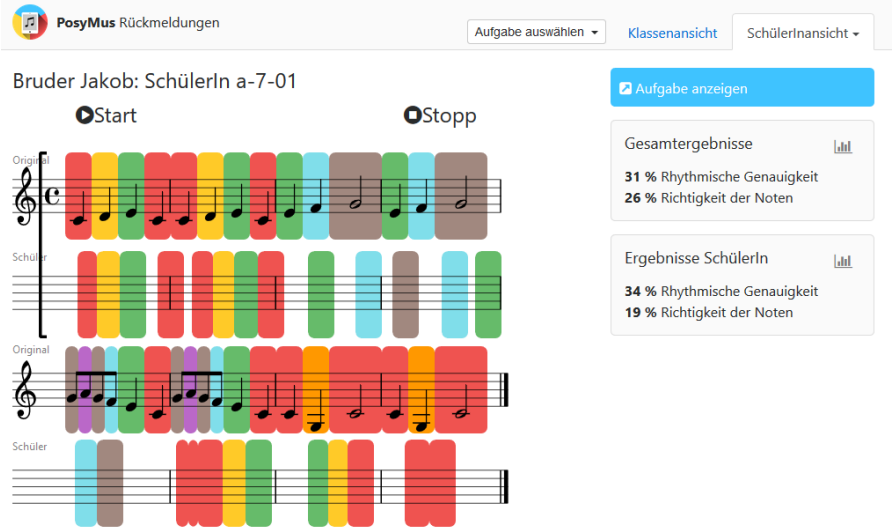
Finken J.: Fig. 5: View 1 and 2



Finken J.: Fig. 6: Class view for contextual items



Finken J.: Fig. 7: Contextual item – Which band plays?



Finken J.: Fig. 9: Single view for practical items

Andreas Lehmann-Wermser, Benjamin Weyel[1]

3. [PosyMus] – the Music Education perspective

Abstract

In this article, we describe our approach for the [PosyMus] digital assessment and feedback system for music class. While our collegues stressed question on informational and technical aspects (see Chapter 2 in this volume), we dealt with questions relevant for music education researchers and teachers.

The two underlying statistically valid sets of tasks were designed to mea-sure music-practical competencies and theoretical and listening skills. We took those tasks and modified them to construct an easy-to-use assessment tool. While testing this tool in schools across northern Germany, we conducted qualitative interviews with the participating teachers to address questions on digital media use in music class in general and for assessment purposes in particular. The analyses of those interviews showed a connection between teacher's personal attitudes towards media technology and their professional beliefs on digital media in teaching/learning contexts. Different types of teacher personalities were found.

Introduction – focus on "competencies" as a new paradigm in German school education

In Germany, music education has reacted somewhat slowly to important shifts of paradigms in educational sciences as well as in policies (Lehmann-Wermser, 2013; 2015). This refers to key terms in the discourse and new

[1] Hannover University of Music Drama and Media

developments like setting up educational standards and can be observed in comparison to other countries as well as other subjects. Among the terms which were largely ignored were competences and feedback. We will focus in our study on the link between competences in music and appropriate feedback.

Contrary to music education, educational sciences have provided a large body of publications that show the importance of feedback for improvement of teaching on all levels (Buhren, 2015; Ditton, 2000; Schratz, 2003; Wollenschläger, 2015; Zierer et.al., 2015). There have been various attempts to implement systematic feedback in schools (Bastian et.al., 2003; Griesel & Gnaudschun, 2014). Especially in teaching foreign languages feedback has been intensively discussed and evaluated (Wilkening 2014).

Feedback has gained attention when the meta-meta-studies of John Hattie were published in a German translation (Hattie, 2009; Hattie, Beywl, & Zierer, 2013):

> *If teachers ask or are at least opening up to students' feedback concerning what they know, what they understand, where they make mistakes, when they understand and when they aren't motivated, then teaching and learning can get in tune and be effective. Feedback can help teachers to make learning visible.* (Hattie 2009, p. 173)

This discourse went by largely unnoticed by music education. There is only one early study of Lösch et al. (2002) that discusses feedback but only in very general terms and is not up to current standards. Thus, very little is known about subject specific ways of feedback in music (Kraemer, 2004).

In terms of assessment it seems especially rewarding to connect with two research projects that developed items suitable for large scale assessment in three of the five fields of behavior ("Verhaltensweisen gegenüber Musik") (Venus, 2001) that are usually listed in German papers: reproduction (singing, playing of given pieces of music), reception (listening and structuring

music), reflection (as knowing about music and understanding music).[2] Two research groups from the University of Bremen (Jordan, 2014; Jordan et al., 2012) and the University of Music Würzburg (Hasselhorn, 2015; Hasselhorn & Lehmann, 2014) developed a large body of items that displayed for several reasons a solid basis for our project. First, the tests were validated and fulfilled the testing standards for advanced IRT-testing. Second, as these research groups proceeded to computer based testing for practical reasons their format was suitable for our purpose. Third, all items were constructed on the basis of a curriculum analysis for all German states, they can be considered ecologically valid in the sense that they cover relevant questions of general music education in lower secondary schools. All in all, these studies provide a solid basis for the study to be reported here that seeks to shed light in the field of every day lessons where curriculum specifications, assessment practices, teaching material and teaching philosophies that should be related to each other mingle at times in hazy constructs and inconsistent philosophies. Systematic feedback could help structure the field. The use of mobile electronic devices seemed an apt mean for supporting this end.

Objectives of the Study

Based on these considerations and the desiderata of music education research the study addressed the following objectives:

1. Constructing an intuitively operable and easy to use integrated testing and feedback system that could be installed on tablets for flexible use in classrooms that should be tested in secondary schools.
2. Gaining knowledge about attitudes of music teachers towards assessment in general and about their statistical competence to understand the results of tests aggregated on class level.

[2] Production (as composing or improvising) and transposition (moving or painting to music) are left aside here.

3. Contributing towards an understanding of ICT use and teachers' specific competences in music lessons.

Schools were selected that offered so called "tablet classes", i.e. classes where every child has either a tablet of his own or sets of tablets are provided for classroom use. Tests were administered consisting of the items adapted from the tests mentioned. More details will be explained below. Meanwhile, the feedback system was developed to inform teachers about the students' results.

The Test

Item base and reconstruction

The items for the testing part of the [PosyMus] e-assessment system were taken (see figure 1) from the validated tests aforementioned: KOMUS (Jordan, 2014) and KOPRA-M (Hasselhorn, 2015). Both projects address specific subjects – the so-called dimensions – deviated from the German middle school music curricula. The KOMUS project addresses the theoretical field of "contextualizing" and the KOPRA-M project the music practical field.

The KOMUS project consists out of four sub dimensions, "verbalization", "contextualization", "listening" and "notation". Verbalization covers knowledge about musical terms, contextualisation includes tasks on the socio-cultural functions of music, listening covers the ability of differentia-ting genres for instance and notations deals with note-reading abilities.

Those subdimensions each have three levels of complexity and for each level there are at least three items. All in all, the researchers of the KOMUS project developed 179 items within the described subdimensions. In order to get all levels of complexity into the [PosyMus] test, one item of each level of difficulty from all four sub dimensions were adopted. Thus, twelve items of the KOMUS item pool were implemented.

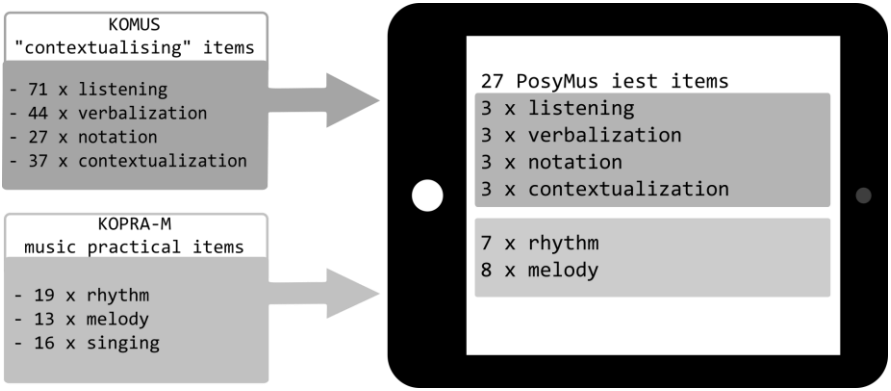


Fig. 1: Sources of posymus item pool.

The original KOPRA-M test is made up of the subdimensions "melody production", "rhythm production" and "singing", and similar to the KOMUS test, each subdimension has items of different difficulties. An overall of 15 items of easy, moderate and hard tasks from the two subdimensions melody and rhythm were adopted for the [PosyMus] test.

Items from the third subdimension – singing – were left out because of the intention of a system, that is able to give feedback without the need of involved experts for processing or rating the data. The processing of a recorded singing voice into musically evaluable digital data is way more difficult than dealing with the digital input values of a virtual melodic or rhythmic instrument.[3]

To get the process chain of the test and feedback system fully automated (see table 1), the items of both, the KOPRA-M and the KOMUS project, had to be transformed in different ways. All practical tasks that were integrated into the [PosyMus] test were modified to work with digital music control data as

[3] For research on automated singing evaluation, see Tsai & Lee 2011, also Molina et al. 2013.

input and output.[4] On the one hand, an input system based on such control data allows an easy construction of interactive multimedia tasks – playing rhythm-patterns to a simultaneously played musical piece, for instance. On the other hand, this control data allows – on the output side – a real-time processing, evaluation and thus, an instant feedback generation.

The two values "pitch" and "time" are being recorded, so the system knows which note (or rhythm sound) is played on which point in time. This approach uses little computing time and bandwidth. The latter point is important, as we distributed the survey (for ease of use) over private and secured local network to devices with comparatively low processing power. Furthermore, control data, other than recorded audio, can easily be processed within the feedback system without human interaction – one main focus of our system.

	Presentation	Input	Output	Processing	Evaluation
KOMUS	Laptop	Laptop	digital data	human	human experts
Kopra-M	Laptop	Tablet & Microphone	digital audio	human	human experts
PosyMus	Tablet	Tablet	digital data	automated	software & teacher

Table 1: Comparison of devices/methods used in the assessment process chain of KOMUS, Kopra-M and [PosyMuS]

[4] One well-known example for an implementation of musical control data is the MIDI protocol.

Rhythm Items

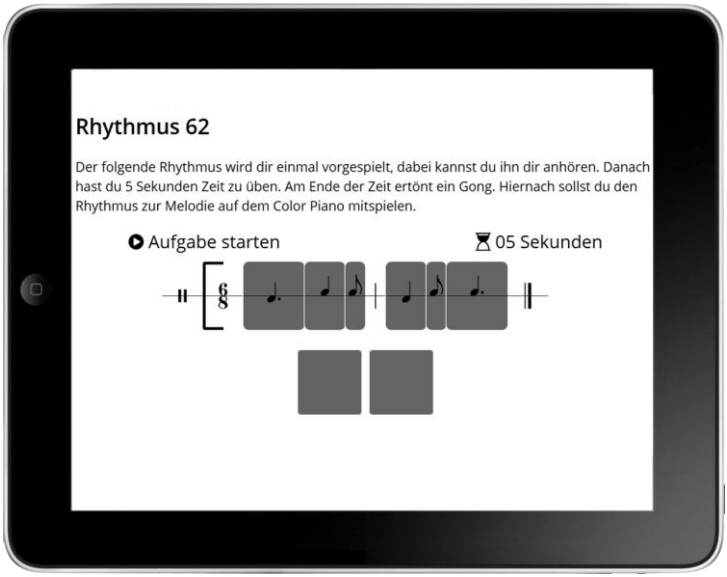


Fig. 2: Rhythm item of the PosyMus test system. The two rectangles below can be hit by touch gestures and represent a high o a low rhythm sound.

Figure 2 gives an example of a task, adopted from Hasselhorn's (2015) KOPRA-M test and modified for the use on mobile devices. Students have to play the notated rhythm along with a musical piece using the two colored buttons on the lower screen area. These can be played like a simple percussion instrument with two discernible pitches. After reading the instructions in the upper screen and listening to the rhythm once, the student has to use both buttons to reenact the rhythm after a short period of practice (in this case, 5 seconds).

What the test system actually records is not the sound itself, it is the two values "note" and "time" for each tap. On the basis of these two information streams, our feedback system is able to visualize rhythmical and tonal correctness. Rhythmical correctness is clearly in the focus of this item type. Obviously, "tonal" correctness can be quite easily achieved here, with only two different tones.

Melody items

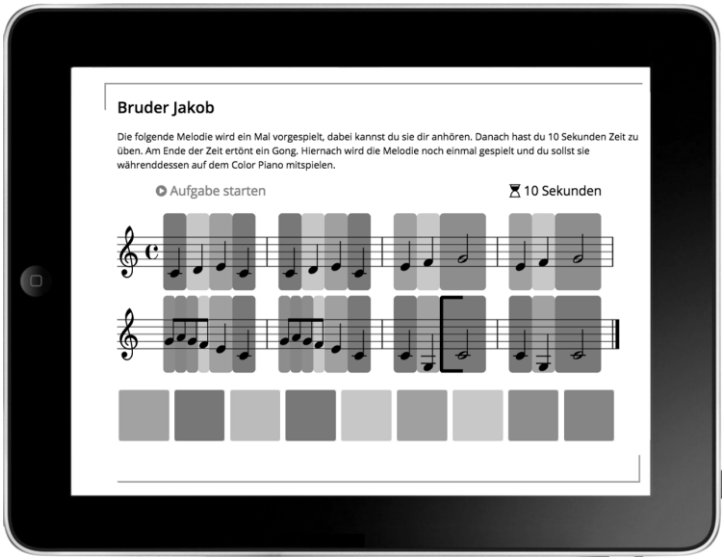


Fig. 3: Melody Item with the color piano in the lower area of the screen. Each rectangle represents one note of a diatonic scale; its color corresponds with the notes above. The rectangles are playable by touch gestures (for a colour version see p. 28).

Melody items are working similar but with a more complex input (see figure 3). The test taker has to reenact a given melody or bass line by using the so

called "Color Piano"[5]. The Color Piano is the input device for items of this type, which intentionally doesn't look like a regular keyboard, so students with private music lessons have less advantages over their classmates. Furthermore, students who are unable to play by score can orientate with the help of the colors. Each color in the piano represents one note from the staff. Like the rhythm items, the melody items produce digital output values describing notes and time with the difference, that there are twelve different value options for "pitch".

"Contextualizing" Items

Most of the items of this dimension are regular text-based multiple-choice or open format items. Some of them where nonetheless complex to implement from a technical point of view; because a good usability was fostered, some of the KOMUS items needed to be made more user friendly. In tasks with visual stimuli, for instance, pictures can be marked directly, rather than have numbers next to them, which have to be inserted into an answer textbox.

The decisions which items to be converted and which not were based on statistical and technical considerations. On the one hand, the different levels of difficulty needed to be kept. This is especially important for the contextualizing items. Whereas in the music practical items different levels of competence can be differentiated within the results of a single item depending on the tonal and rhythmical correctness, this isn't possible in the "contextualization" items since many of their answers can either be right or wrong. Often, there is no other i.e. non-dichotomic result that could allow a detailed differentiation. Thus, the different difficulty levels within all sub dimension are crucial and needed to be retained.

[5] The color piano was developed and empirically tested as an input device for this kind of items within the KOPRA-M research project, see Hasselhorn 2015, pp. 54ff. In order to fit the color piano on one screen with the instructions and the notation of the melody, we had to slightly modify it to a more horizontal orientation.

On the other hand, several technical limitations needed to be circumvented. Singing, as described above, is hard to evaluate in an automatic process, that whole sub dimension was dropped. Beyond that, the approach of an easy-to-use interface that works with standard consumer hardware, needed to be economical in terms of transfer data amounts. Both, the "contextualizing" and the music practical items need and produce lots of data that need to be sent and received over wireless local area network. In order not to overcharge the computing time and bandwidth of standard network hardware, items with too huge (respectively long) audio and visual stimuli were dropped. However, some items with large tonal or visual stimuli could be kept, because we were able minimize the required data rates, for instance by using efficient data reducing methods like downsampling and audio compression. At the end, 27 items were migrated into the [PosyMus] test.

Testing Process

So far, the test was administered in twelve classes from six middle schools across northern Germany. For each test session, 18 items were selected in order not to exceed regular 90-minute lesson: 45 minutes for the test itself, another 45 to set up equipment, give instructions, and get familiar with the gear and practice. Sets of items were changed from class to class.

For this first period of testing, the necessary network- and test-hardware was brought by the researchers. It consists of a home Wi-Fi router and a low-cost laptop running the server software.

The Feedback

Some problems need to be addressed that are either of a more technological nature or related to the questions of teachers' use of ICT. For the field mentioned first the feedback design for the music practical tasks is of interest. As for rhythms and melodies for each note a value for time and pitch can be recorded and analyzed. In accordance with findings from music psychology

(i.e. McKinney & Moelants, 2006) a time slot beyond the numerically correct onset was defined in which every played note counts as “on time”.

Yet the question remains what kind of output teachers want and need and are able to understand. For instance a summative output indicating the individual overall percentage of correct rhythm and pitch would give hints on achievement aggregated on a class level. However, it would not help to identify patterns of mistakes, difficult bars and the like. In this case it would be desirable to have a dynamic feedback that shows performance over time, which fits the time-dependent character of musical behaviour.

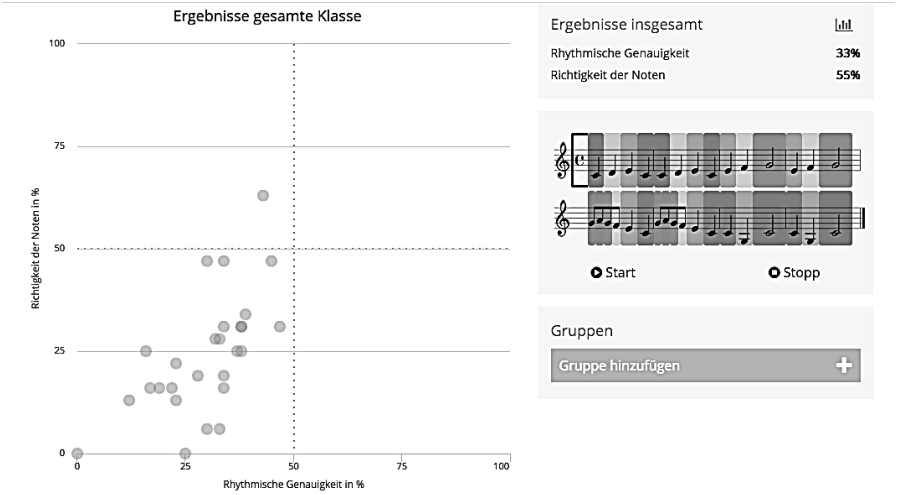


Fig. 4: Screenshot example of a feedback on playing a melodic line showing the results of the whole class. The axes resemble rhythmic and tonal correctness with percentage index up to that point. Perfect performance would be showing in the upper right-hand corner. Each dot on the left side represents a student. On the right side of the screen the score of the music is shown with several options for levels of feedback.

To make those processes visible, a dynamic system was developed that shows the average percent of rhythmic and tonal correctness up to an arbitrary point in time. Normally a student would start in the upper right-hand corner as it is easy in the beginning to synchronize and to foresee difficult bars. Usually there might be a tendency of some students migrating in the chart to towards the lower left quadrant, which indicates a lower performance. This might be dependent on increasing difficulty of certain bars. Accordingly, one might expect students to perform less well at the beginning of the second staff (see figure 4).

The right half of the screen offers various modes of feedback: output can be generated for individual students or for the whole group. Results can be matched with other current results from the particular student or series from earlier periods. The depth of output needed is part of the research. Once data are generated the degree of detail is not a technical problem but rather a question of practical use and understanding on the teachers' side.

The whole system is based upon open source software and the code written by the [PosyMus] team will also be open source, so everyone with web development skills can reuse it. For a detailed technical description, see BREITER et al 2018 in this publication.

Qualitative study

The [PosyMus] project not only included the development and testing of an assessment tool in schools, but to gain information about how music teachers think about digital media and e-assessment in general. Our overall goal was to gain knowledge on possible teacher-side factors for a successful implementation of a digital test and feedback system in music lessons. Therefore, questions of acceptance, usability and statistical literacy are presumed to play a major role.

Structured Interviews are held and analyzed according to grounded theory methodology (Straus & Corbin, 1996). The sample included those teachers,

who took part in the test- and feedback evaluation, and additionally, teachers that didn't participate were interviewed as well, to get some more contrasting views. The focus of these interviews lies on teachers' beliefs on e-assessment and ICT in music lessons, especially its assumed usefulness.

Teachers can image a broad variety of advantages of an e-assessment system for music class. One aspect concerns the rapidness of electronic competency measurement and feedback:

> *Well, I think that digital media is really helpful in the field of diagnose, because i can let them do the input within a test instantly. And the software can evaluate these inputs instantly regarding musical competencies and maybe give some sort of protocol instantly to the teacher, so that the teacher can, at best, get an instant overview over the musical competencies.* (secondary school teacher A, from Lower Saxony)

This quotation shows that this teacher from Lower Saxony is well aware of the possibilities of an instant feedback, which can be useful for his teaching conception. An overview of the competencies of his students, measured at the beginning of could bring – in his view – advantages in planning and conception of his didactic work.

In contrast to that, Teacher B from Lower Saxony sees no advantage in e-assessment systems for the subject of music, mainly because assessment in general isn't important in music class:

> *I rather trust my gut feeling when assessing students [...] assessing in general is rather secondary in the subject of music.* (secondary school teacher B from Lower Saxony)

Assumingly, this teacher isn't willing to use any kind of e-assessment tool in music class. In the course of the interview, she describes her idea of good music lessons, which leans mainly on aesthetic experience rather than measurable competencies.

Connections Between Private and Professional Attitudes

Further analyses also revealed a strong connection between affinities to technology in private surroundings and attitudes towards ICT as a useful medium in schools. Not surprisingly, those interviewed teachers that describe themselves as "tech-savvy" have more positive attitudes towards ICT-use, but also towards e-assessment in schools. And, with no exception, every teacher, that has such positive attitudes, reported, that he or she had been started to use ICT for private reasons even before his or her carrier as a teacher begun. Thus, they seem to adopt their technical affinities from private to their professional selves. Figure 5 is based upon the statements of one tech-savvy teacher and illustrates that reciprocal process of transferring competencies and attitudes from private to profession. Not mentioned in that figure are several factors that assumingly have an influence on attitudes and motivation, such as hardware provisions in school or supportive behaviour from administrative side. So far, it is also unclear if and how attitudes towards electronic assessment techniques and devices are linked. The investigation goes on. But it seems interesting, that the origin of a positive attitude for whatsoever ICT-use derives from private circumstances. Further investigation will also hopefully answer the question, which role the academic teacher training can play on cultivating positive (and maybe negative) attitudes towards ICT and e-assessment.

All in all, the interviewed music teachers seem to use digital media mainly as tools for music reception. Only few of them reported a use for musically creative or productive tasks. Examples for mentioned applications can mostly be categorized into presentation software (like Keynote), informational websites.

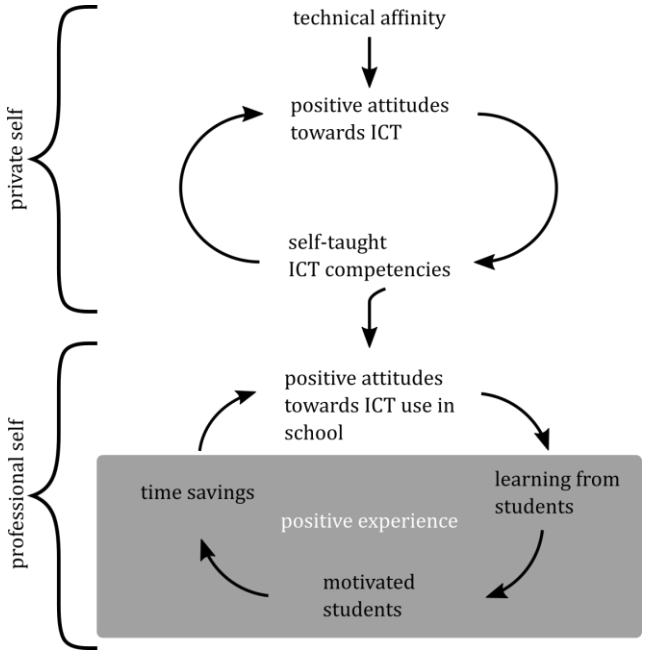


Fig. 5: Reciprocal model of the evolvement of positive attitudes towards ICT use in class, as reported by music teacher C.

Implementing E-assessment in music class

Eickelmann (2010) highlighted sets of inhibiting (p. 281) and supporting factors (p. 284) for an implementation of information and communication technology (ICT) in schools. Some of them can be found in our interviews as well. Negative beliefs and attitudes towards ICT we found often dealt with ineffectiveness: teachers mention an assumed time-consuming quality of ICT-use in music lessons. Furthermore, the lack of ICT literacy on either teacher or student-side is one of the inhibiting factors that teachers see.

Insufficient ICT-provisions and the lack of supportive structures in schools seem also problematic.

Factors of a successful implementation are positive attitudes towards ICT, appreciation and support from school administration, colleagues and parents and a working and well-maintained technical infrastructure in schools.

Implementing innovation in general and especially ICT in schools is a somewhat complicated process with several influencing parameters on teachers', schools' and governmental levels (Fullan 2001). Our study revealed some factors on teachers' side – personal attitudes and experiences, mainly derived from outside of school and teacher training seem to be of major importance. On the other hand, our practical in-school-testing showed that some technical and organizational requirements need to be fulfilled (see Breiter at al. in this publication). The investigation on how to build an effective electronic feedback-systems for music teachers has shown that it is important to give individual and time-discrete information of students' performances in music practical tasks. This way teachers can transform the given data into classroom intervention and lesson plans.

References

Bastian, J., Combe, A., & Langer, R. (2003). *Feedbackmethoden: Erprobte Konzepte, evaluierte Erfahrungen. Praxis* [Feedback methods: Proven concepts, evaluated experiences, practice]. Weinheim: Beltz.

Buhren, C. G. (2015). Feedback – Definitionen und Differenzierungen [Feedback – definitions and differentiations]. In C. G. Buhren (Ed.), *Pädagogik. Handbuch Feedback in der Schule* [Pedagogy. Handbook of feedback in school] (pp. 11–30). Weinheim: Beltz.

Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht [Quality control and quality assurance in school and class]. *Zeitschrift für Pädagogik, 41*, 73–92.

Eickelmann, B. (2010). *Digitale Medien in Schule und Unterricht erfolgreich implementieren* (Empirische Erziehungswissenschaft, 19) [Successfully implementing digital media in school and class (Empirical educational science, 19)] (pp. 311–329). Münster: Waxmann.

Fullan, M. G. (2001). *The new meaning of educational change*. New York: Teachers College Press.

Griesel, P. & Gnaudschun, H. (2014). Feedback-Verfahren im Unterricht einführen: Kommentierte Beispiel aus einem Baukastensystem [Introducing feedback in class: commented example from a modular system]. *Pädagogik, 66*(4), 10–13.

Hasselhorn, J. (2015). *Messbarkeit musikpraktischer Kompetenzen von Schülerinnen und Schülern: Entwicklung und empirische Validierung eines Kompetenzmodells* [Measurability of students' music practical competencies: Development and empirical validation of a competency model]. Münster: Waxmann.

Hasselhorn, J. & Lehmann, A. C. (2014). Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz [The development

of an empirically verifiable model of music practical competency]. In B. Clausen (Ed.), *Musikpädagogik Forschung und Lehre, 36* (79–95). Münster: Waxmann

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement (Reprinted).* London: Routledge.

Hattie, J. N., Beywl, W. & Zierer, K. (2013). *Lernen sichtbar machen: Überarbeitete deutschsprachige Ausgabe von "Visible learning"*. [Visible learning: edited German- language edition] Baltmannsweiler: Schneider Hohengehren.

Jordan, A.-K. (2014). *Empirische Validierung eines Kompetenzmodells für das Fach Musik: Teilkompetenz "Musik wahrnehmen und kontextualisieren"* (Empirische Erziehungswissenschaft, 43). [Empirical validation of a competency model for the subject of music: sub-competency „perceiving and contextualising music" (Empirical educational science, 43)], Münster: Waxmann.

Jordan, A.-K., Knigge, J., Lehmann, A. C., Niessen, A. & Lehmann-Wermser, A. (2012). Entwicklung und Validierung eines Kompetenzmodells im Fach Musik - Wahrnehmen und Kontextualisieren von Musik [Development and validation of a competency model for the subject of music – perceiving and contextualisation of music]. *Zeitschrift für Pädagogik* [Journal of pedagogy], *58*(4), 500–521.

Kraemer, R.-D. (2004). *Musikpädagogik - eine Einführung in das Studium* [music education – an introduction into academic studies]. Augsburg: Wißner.

Kultusministerkonferenz (2004). *Bildungsstandards der Kultusministerkonferenz* [Educational standards of the Kultusministerkonferenz]. Retrieved from: https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html [15.08.2020].

Lehmann-Wermser, A. (2013). Music Education in Germany: On Politics and Rhetoric. *Arts Education Policy Review*, *114*(3), 126–134.

Lehmann-Wermser, A. (2015). Teaching Music in Germany. In S. Figueiredo, J. Soares & Schambeck R. F. (Eds.), *The Preparation of music teachers: A global perspective* (pp. 219-240). Porto Alegre: ANPPOM.

Lösch, C., Merz, U. & Kivi, A. (2002). Evaluation und Feedback für Lehrende und Lernende [Evaluation and feedback for teachers and students]. *Diskussion Musikpädagogik*, *16*, 18–24.

McKinney, M. F. & Moelants, D. (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels? *Music Perception*, *24*(2), 155–166.

Molina, E., Barbancho I., Gómez, E., Barbancho A.M. & Tardón L.J. (2013). Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 744–748.

Preece, J., Rogers, Y. & Sharp, H. (2007). *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons.

Ras, E., Swietlik, J., Plichart, P. & Latour, T. (2010). TAO – A Versatile and Open Platform for Technology-Based Assessment. In M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt & V. Dimitrova (Eds.), *Sustaining TEL: From Innovation to Learning and Practice* (pp. 644–649). Berlin: Springer.

Schratz, M. (2003). *Qualität sichern: Schulprogramme entwickeln* [Assuring quality: developing school mission statements]. Seelze: Kallmeyer

Shneiderman, B. & Plaisant, C. (2005). *Designing the user interface: Strategies for effective human-computer interaction*. Boston: Pearson.

Strauss, A. L. & Corbin, J. (1996). *Grounded Theory – Grundlagen qualitativer Sozialforschung* [Basics of qualitative research]. Weinheim: Beltz.

Tsai, W. H. & Lee, H. C. (2011). An automated singing evaluation method for Karaoke systems. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2428–2431.

Venus, D. (2001/1969). *Unterweisung im Musikhören* (4. Ed.). (Musikpädagogische Bibliothek 30) [Instruction on music listening (4. Ed.) (music educational library 30)]. Wilhelmshaven: Noetzel.

Wiliam, D. (2011). What is assessment for learning? In *Studies in Educational Evaluation, 37*(1), 3–14.

Wilkening, M. (2014). *Ein Überblick über Aufsätze und Praxismaterialien zum Thema Feedback* [An overview of articles and practical material on the topic of feedback]. Retrieved from: http://web.fhnw.ch/plattformen/hattie-wiki/w/images/Literatur_Feedback_mwi.pdf. [DATUM].

Wollenschläger, M. (2015). Schülerfeedback: Forschungsergebnisse und Empfehlungen [Feedback for students: research findings and recommendations]. In C. G. Buhren (Ed.), *Pädagogik. Handbuch Feedback in der Schule [Pedagogy. Handbook on feedback in school]* (pp. 263–283). Weinheim: Beltz.

Zierer, K., Busse, V., Wernke, S. & Otterspeer, L. (2015). Feedback in der Schule – Forschungsergebnisse [Feedback in school – research findings]. In C. G. Buhren (Ed.), *Pädagogik. Handbuch Feedback in der Schule* [Pedagogy. Handbook on feedback in school], (pp. 31–50). Weinheim: Beltz.

Johannes Hasselhorn[1] & Jens Knigge[2]

4. Technology Based Competency Assessment in Music Education: The KOPRA-M and KoMus Tests

In this chapter we'd like to summarize the research done in the last ten years regarding assessment and modelling of competencies in German music education. Since this research is closely connected to issues of technology-based competency assessment (tbca) this topic will be the main focus. We use the KOPRA-M and KoMus tests to illustrate what tbca can look like to show the main advantages and to discuss challenges and further perspectives.

Competency-orientation in the German school system

There is a long tradition of interest in the evaluation of music ability in music education and music psychology. There are tests of musical aptitude, of appreciation and attitude, of musical performance, and of musical achievement.[3] However, a review of this research shows that the focus is primarily on aptitude and performance rather than on tests of musical achievement (Colwell, 1999). This is also the case in Germany, so the discussion surrounding assessment of competencies has been hitting German music education unprepared. This discussion has been sparked off by changes in the educational system since the beginning of the century. The ultimate goal of this reform process was the implementation of national standards ('2004') whereby the primary focus was directed from the input to the output of the

[1] Technical University Dortmund

[2] Nord University Levanger (Norway)

[3] For historical information about music ability evaluation see, e.g., Boyle, 1992; Kormann, 2005.

system.[4] The development of national educational stan-dards in Germany was a cooperative project of the federal government, local school boards, and educational researchers. Some important tasks were assigned to researchers in the field of education, first of all the development of competency models and corresponding assessment instruments, whereby the central goal is the empirical measurement of pupils' abilities and skills (competencies) at certain points of their school career (e.g. Hartig et.al., 2008).

Although national standards have not been implemented for the school subject music, all music curricula as well as many teaching methods/concepts and books are strongly influenced by the idea of competency-orientation at present (Knigge, 2014). This has been problematic to some extent, given that German music education had no tradition in neither testing nor modelling competencies. This led to the unsatisfying situation, that all music curricula became output-oriented, while the validity of the formulated competency standards, their content, their structure, and last but not least their measurability (in terms of competency assessment) were questionable. Against this background some research studies have been conducted to address the topics of modelling and assessing musical competencies.

Competency studies: current state of research

The first study which started in 2007 was an exploratory study and was entitled "KoMus" (KoMus stands for "competency model for music"). One of the first tasks of project KoMus was to identify competency domains and subdomains in music that were theoretically convincing and practically relevant.

[4] For detailed information about the broad and complex conception of the national standards and the related reforms in German educational system see Klieme & Maag Merki, 2008.

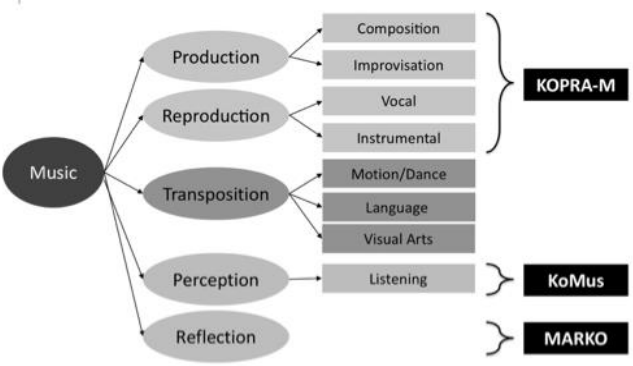


Fig. 1: Competency domains in (German) music education and corresponding research projects.

Figure 1 shows a suggestion made by Dankmar Venus as early as in the late 1960s. He proposed five domains or as he called them "patterns of musical behaviour": production, reproduction, transformation, perception, and reflection. Supplemented by an analysis of German and international music curricula, we came to the conclusion that there is a strong national (and international) consensus for at least the four dimensions: production, reproduction, perception, and reflection, so they could serve as a basis for model and test development (Niessen et.al., 2008).

On this basis the researchers in the project KoMus developed a competency model and a corresponding test for the domain of music perception, finally published in 2012 (Jordan et.al., 2012). From 2010 to 2013, a following study (KOPRA-M) focused on the domains of production and reproduction (Hasselhorn, 2015). Lateron started the project MARKO, which deals with the reflection about music or namely the "music related aesthetic argumentation" (Knörzer et al., 2015).

Why tbca is relevant to competency assessment in music education

Before we discuss the KOPRA-M and KoMus tests in more detail, it seems appropriate to clarify the relation between our studies and the topic of the symposium documented in this book. Or in other words: Why is the approach of technology-based competency assessment relevant to test development in music education?

Honestly, we had no tbca in mind when we started with KoMus in 2007. Even more honest: We developed test items for almost a year on a paper-and-pencil basis, tried these items in classrooms with 20 to 30 students and played audio files over a speaker system standing in front of the class. The problems we faced in this paper-and-pencil setup are the following (Knigge, 2010):

- shared audio unit corrupts test fairness (different distances to the audio unit, noise from outside, etc.)
- test-tempo is the same for all test-takers
- especially when it comes to critical reflections/aesthetical judgments individual handling of the recordings is necessary (e.g. audio files had to be played repeatedly and in individual time frames)
- copying off of a neighbour ("cheating")
- rating of productive tasks is not possible in group testing

In sum, the results we got working with this setup, led us – or better: forced us – into the world of tbca. Actually, tbca helped us to resolve the mentioned problems, something we'll discuss later in more detail.

Although tbca helps on the one hand, it creates challenges on the other hand. Almost all solutions use network technology. And in the context of distributing test items and collecting data on a network basis, you'll face at least some of the following problems:

- file size (especially in web-based settings)

- sound quality
- latency
- coherence between media types (sound & video files), assessment software, and hardware (e.g. no school has the same PC, OS, data transfer rate, browser, media player etc.)
- scoring of open-ended items and productive/creative tasks.[5]

The KoMus competency model and test

In the next section we want to illustrate the mentioned challenges and – of course – the solutions we found in the projects. First, we introduce the competence model, the test is based on. Then, we'll continue with some information about the test setup, and finally invite the reader to try some test items.

Competency model "perceiving and contextualizing music"

When you look into German music curricula – something we did very intensively ahead of the model development – you'll see that a model for the domain of perception has to include not only listening but the interaction between listening and knowledge about the social and historical contexts of music, the use of music notation, and the use of a technical vocabulary. We therefore named the model "perceiving and contextualizing music" using the dimension of perception as a "main"-dimension, and the others as three subdimensions (see figure 2; for more details see Jordan & Knigge, 2010; Jordan et al., 2012).

[5] Editor's note: The project was conducted in 2008, this chapter was written in 2017. The situation concerning hardware and internet access in schools has improved somewhat since.

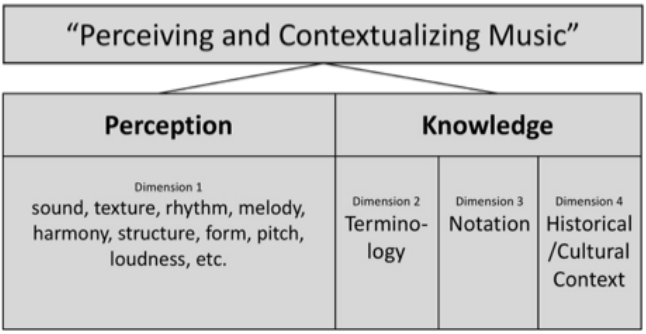


Fig. 1: Schematic version of the KoMus competency model.

It is important to note that the subdimensions are of interest only in their relation to perception. We do not assess these competency facets in isolation (e.g. score-reading items) but in conjunction with listening tasks. Consequently, the KoMus model contains a complex competency structure.[6] This complex structure needs to be addressed on test and item level as well.

Competency test

Seen from a technical perspective, the test has an individual setup, where every student works on his or her own laptop and with a set of headphones. The test itself runs on ILIAS or Moodle platforms and is distributed through a data network.

In the beginning, we took an Ubuntu server to every classroom and connected all clients with LAN cables to the server. Since the internet connections of schools have been improved significantly over the last years, we do not use a local server anymore, but a web-based installation of Moodle.

[6] While figure 1 only shows the dimensional structure, a translated and more detailed version of the validated model is included in the appendix section. There you can see the content of the model. Each dimension has several levels, and each level describes in detail the competencies students have when they reach a certain level.

Since 2015, the KoMus test exists in two versions: the original long version has 83 items, takes about 90 minutes and has satisfying reliability (Jordan et al., 2012). Two years ago, we tried a short version which has only 27 items and a corresponding shorter test time. Overall Reliability is still satisfying, while the first and last dimensions are less reliable (Harnischmacher & Knigge, 2017).

	KoMus_long	KoMus_short
Number of Items	83	27
Reliability	.81 {.82, .81, .79, .69}	.77 {.62, .77, .71, .60}
Duration	90 min.	30 min.

Table 1: Test versions and their main features. In parenthesis are the reliability coefficients for the subdimensions

Regarding psychometrics, it's worth mentioning that we used item response theory or more specific: a generalised multidimensional Rasch model for all analyses and validation procedures (Wu et.al., 1998). Analyses show that the item set fits the Rasch model very well (Jordan et al., 2012; Knigge, 2010). Additionally, the test has a very continuous distribution of the items in relation to the difficulty scale: we have very difficult items as well as medium and easy items. Figure 3 (so called "wright map"; Wilson, 2005) shows the distribution of the items over the competency model (test takers are symbolised as crosses, items as squares).

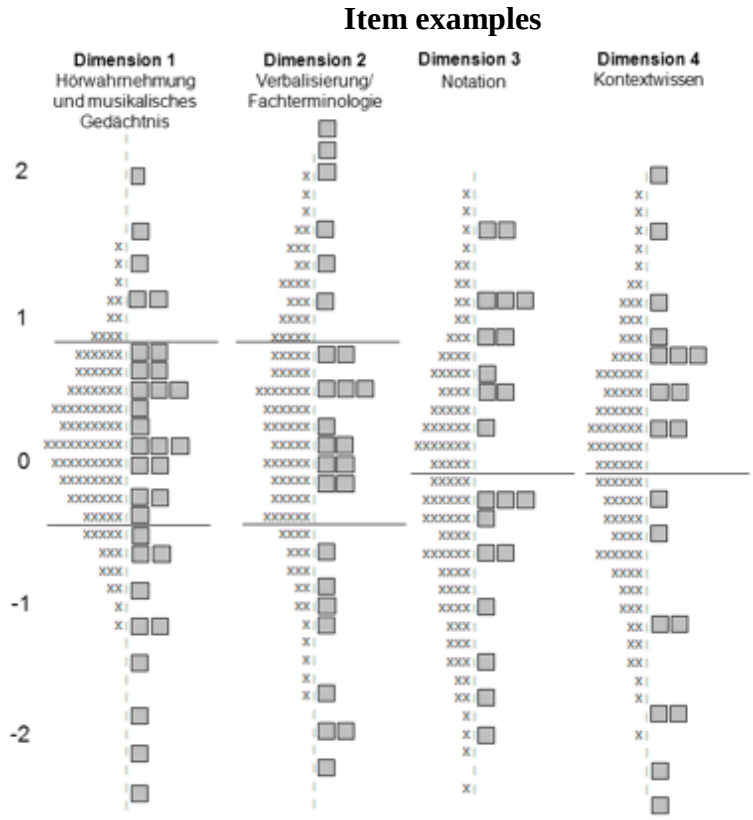


Fig. 2: KoMus test wright-map (Jordan et al., 2012). Items and test takers illustrated on the same metric; the vertical axis represents item difficulty/persons competency (unit: logit). The dimensions 1 through 4 refer to the ones mentioned in figure 1.

The item set consists of open-response, short constructed-response and closed-ended questions. The stimuli include a very broad range of styles and genres (jazz, pop, rock, hip-hop, 'classical', gospel, world music, etc.). During the test the test taker is asked to read the task description first. Afterwards he/she could press the play-button to start the audio stimuli. Sometimes additional visual material (pictures, scores etc.) is provided. Finally, the test

taker can respond to the task by choosing a given answer (e.g. forced choice item) or writing a short text (open-response).

Instead of printing several item examples and describing them in detail with words (but without the sound files), we decided to invite the reader to try out some items by themselves: http://jensknigge.info/site/komus/

The KOPRA-M model

A second example of competency modelling in the field of music is the KOPRA-M model. It claims to model music performance competence. Although the measurement of music performance has a long tradition (see West, 1992; Helms, 2005), this model combines both a structural model of cognitive competence and a group test setting (Hasselhorn, 2015).

The process of modelling the structure of music performance competence was based on the theoretical ideas of Weinert (2001) and understands competence as domain-specific, context-related, cognitive performance disposition (Klieme et al., 2003). This is consistent with the cognitive model of music performance by Lehmann and Ericsson (1997) which consists of the three steps goal setting, motor implementation, and self-evaluation. In a competence view, all three steps are necessary for a high level of music performance competence.

The first aim in the modelling process was to identify possible, separable dimensions of a music performance competence, which means to find and describe content-based and in their cognitive requirement different classes of domain-relevant demand situations. Therefore, several models and results from music psychology (Boyle & Radocy, 1987; McPherson, 1995) and music pedagogy (Elliot, 1995; Jank, 2009; Venus, 1969) were analyzed and compared with German curricula (Hasselhorn & Lehmann, 2014). At the end of this process a three-dimensional structural model of music performance competence was identified.

The *singing* dimension includes singing of melodies and accompanying parts. The second dimension is *playing instruments* which is very similar to singing. But instead of singing melodies and accompanying parts should be played on instruments. In the third dimension, *playing rhythms*, students have to play different rhythmic pattern in different tempos (Hasselhorn, 2015, pp. 39–43).

Because this is a cognitive model and not a didactic model, some in music lessons really famous demand situations are missing in the model: Transposition tasks and creativity tasks. Transposition which means transferring music in another kind of art or the other way around, seems to be a special kind of task which is already a part of the three defined dimensions. Regarding creativity there is a lack of basic research in the field of music as well as in general. On the one hand, we do not have a psychometric good fitting test of musical creativity (Hickey, 2001; Hickey, 2013). On the other hand, there is still a discussion about the generality-specificity question regarding creativity (Barbot et.al., 2016; Baer & Kaufmann, 2005; Lothwesen & Lehmann, 2017). That's why it seems quite possible that a fourth creativity dimension might be added in future. But it is also possible that creativity might be identified as a part of each of the three existing dimensions.

Technical group test setup

Aim of the KOPRA-M project was to develop a test to measure music performance competence in a group setting to be able to conduct this test in an economically useful way. This aim was a reaction to the common way of music performance measurement where usually one student performs in front of a jury of three to seven judges (Hasselhorn & Wolf, 2017). This quite time consuming and expensive setting has led to a situation where music performance measurements have been cancelled, for example in the NAEP studies (Keiper et al., 2009). So, the purpose was to construct a TBA system to be able to test a whole class at the same time.

The idea was to record students singing and playing to judge it later. Therefore, in the KOPRA-M test each student is sitting at an individual workplace. It consists of

- a laptop on which current task is presented including text or notesheets if this is in the task,
- a headset to be able to listen to music from the laptop and to record the voice in the singing tasks,
- a tablet-pc with the CMG-application (Hasselhorn & Grollmisch, 2014) as digital interface for playing instruments and rhythms, and
- a divider for acoustic separation.

All student workplaces are connected with a central PC in a local network. A teacher can provide tasks through a JAVA-based browser application for simultaneous task processing. Each student is working on the same task at the same time. This is very important for the singing tasks due to minimizing acoustic irritation. All students' answers (singing and playing) are recorded and saved on central PC.

Students have not to answer typical test items but have to act in dimension specific demand situations. Of course, these situations are organized in tasks. In *singing dimension*, tasks consist of a play along with or without the voice, students should sing. In some tasks students are allowed to listen to a known or unknown song, in some they are not. They have a short time to rehearse. After that time, the play along starts with playing the starting tone, and the setup is recording what students are singing. In *playing instruments* dimension, students have to play melodies or accompaniments to play alongs. It is nearly the same than singing tasks. As a substitute for a conventional instrument, which is not useble for a group test setup, the Colored music grid (CMG) app is used (Hasselhorn & Grollmisch, 2014). In this app, students can play twelve different diatonic tones in C major by pressing twelve different coloured buttons. These colours are also presented in the notesheets on the laptops. A validation study showed that the musical results students

produced with the CMG app are equal to what they produce with a piano (Hasselhorn, 2015, pp. 100–109). In *playing rhythms* dimension, tasks are slightly different. Here, students have to repeat one- or two-bar patterns matching to an 8-bar recording. Some patterns are only shown before playing, some are played, and some are both played and shown. Recordings vary in tempo.

Data analysis

Rating	Description
1	Child is nearly or totally accurate singer
2	Child sings with some accuracy, beginning in the established key
3	Child sings song with some accuracy, starting in a different key than established, or modulates within the song
4	Child sings/chants melodic shape at significantly different pitch
5	Child sings/chants song with a different melodic contour than the song
6	Child does not even try to sing the correct song OR system error (nothing to hear)

Table 2: Extended singing performance assessment rubric (Hornbach & Taggart, 2005)

For statistical data analysis the recordings had to be translated into ratings. Therefore, an already existing singing performance assessing rubric (Hornbach & Taggart, 2005) was used and extended (see table 2). For judgments in instrumental and rhythmic dimension new assessing rubrics were created

analogous to the Hornbach & Taggart (2005) rubric. Each of more than 20.000 recording from 445 form 9 students was judged by at least two expert raters. An intraclass correlation ICC (1,1) = 0.80 showed a good interrater reliability.

With these data, an analysis was run to evaluate if the hypothesized structure of three separable dimensions represents data well enough. Using the item response theory in general and a partial credit model in special, the three dimensions model showed much better fit indices than any other possible model (Hasselhorn, 2015, pp. 136–140).

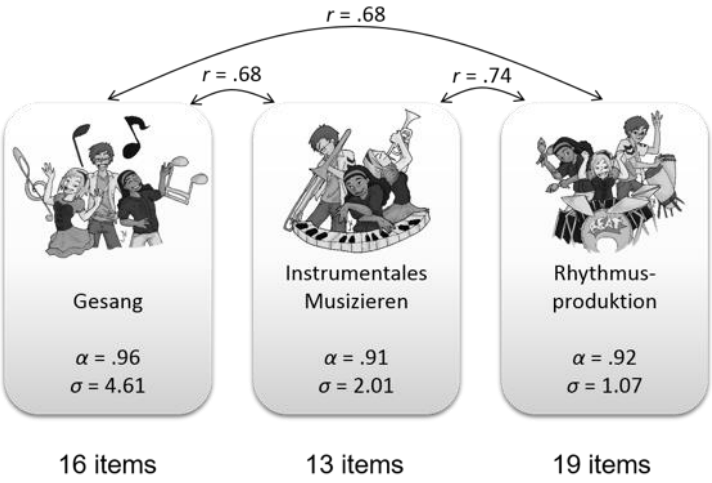


Fig. 4: Structural model of music performance competence (KOPRA-M).

The final test does fit to the structural model and has very good values for the IRT common quality criteria (see figure 4). The test includes a total of 48 items.

TBCA in music: first results

Research on competence in music education is still in its infancy. Nevertheless, some first results can be reported. The first and maybe most important result is: It is possible to model and measure musical competencies in a valid and reliable way. For both, the KoMus and the KOPRA-M model we do have a data-based and detailed description of students' competencies for grades six and nine.

Another result is the data-based evidence for the implicitly assumed huge variance of musical competencies in school. A much wider distribution than the normal distribution was found for singing (four times larger than expected in grade 9) and instrumental competencies (six times larger than expected in grade 9; Hasselhorn & Lehmann, 2015). Although teachers assume this from their individual experience, we now have empirical evidence and a first impression about how large music performance heterogeneity in school actually is.

Further findings suggest the importance of certain factors for competence development. Regarding perception competency (KoMus) students in higher secondary schools (Gymnasium) showed 0.41 to 0.68 SD better performances than students in lower secondary schools (Haupt-, Real-, and Gesamtschule) (Jordan, 2014, p. 166). Girls performed up to 0.40 SD better than boys (p. 163), students with a higher socioeconomic status 0.32 to 0.45 SD better than students with a lower SES (p. 165). Playing an instrument/singing out of school (e.g. private music lessons, singing in a choir) generated an advantage of 0.16 to 0.29 SD (p. 163). Regarding music performance competencies (KOPRA-M) girls scored much higher than boys (0.41 to 1.01 SD), students in higher secondary schools (Gymnasium) much higher than others (0.81 to 0.96 SD), playing an instrument with private lessons gave a huge advantage (1.01 to 1.17 SD), and a special musical training in early childhood caused a slight advantage (0.12 to 0.15) (Hasselhorn & McElvany, 2016).

Especially in educational psychology, the correlation between motivation and competence development is a central research area. In music education research we just started to use the developed competence tests to investigate this correlation for musical learning. Harnischmacher and Knigge (2017) found a correlation between motivation of music-related action and perception competence (KoMus-test) of $r = 0.34$ in a sample of 600 grade 7 students. Together with the family's interest in music and musical practice (out of school), motivation explains 20 % of the variance in student's competencies.

In an intervention study to foster perception competence by motivation (Hasselhorn et.al., 2017) 236 grade six students got two music listening lessons. In one group, students were taught considering the principles of self-determination theory (Ryan, & Deci, 2017). Motivation was significantly higher in this group in a post-test ($F = 5.78$, $p < .01$, $\eta^2 = .05$). The KoMus score was equal in both groups in the post-test but significantly higher in the motivation group 8 weeks later ($F = 4.76$, $p < .01$, $\eta^2 = .04$), which underlines the long-term effect of motivation in context of learning.

TBCA in music: potential for future work

Technology based competence assessments like the KoMus and the KOPRA-M tests could be used in future research to conduct a wide spread of evaluation studies, intervention studies, and experimental studies. These tests might help us to understand what works in contexts of music learning and how, when, and why it works. Evaluation studies could identify the potential of certain teaching concepts. Intervention studies might address the impact of special problem driven actions. And experiments have the potential to quantify isolated psychological learning mechanisms.

Furthermore, there are also school settings in which such tests could be used potentially. Teachers might use them to evaluate a series of lessons. They even could use them in a diagnostic way in the beginning of a schoolyear to

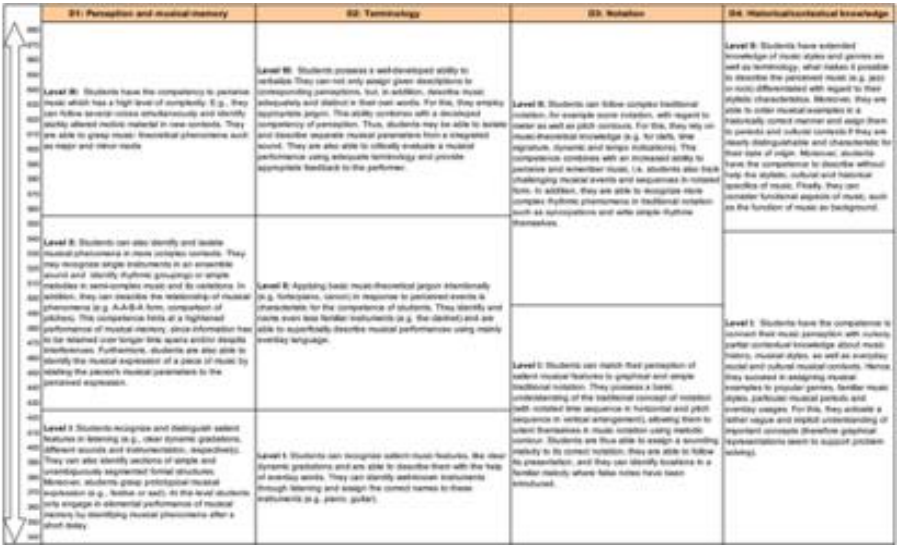
get more information about the class and thereby improving their concept of teaching. And even in the absence of a teacher, the tests could be used as an individual feedback system for students in form of evaluation, self-evaluation, and learning management systems. In this way, students would get an objective source of information for their own learning process additional to teachers', peers', and parents' feedback.

To be able to realize these ideas in the future, a lot of work has to be done. Firstly, there is a need for validation studies with different age samples. At the moment, the KoMus test is validated for grade six students, the KOPRA-M test for grade nine students. Much more data is needed to enlarge this. Similarly, we need more data for stable standardization, and thereby a reliable norm sample. This is a requirement to classify single results in a secure way later on.

To improve the economical use of the tests, much work has to be done to make the test systems run in all technical situations in schools. Browser-based programmes should be a useful solution for this. Additionally, an automatic data-analysis tool is needed. For the KOPRA-M test self-learning algorithms might replace the expert ratings. Last but not least, we need a good documentation and trainings for teachers which help them understanding the results of the tests. In the end, computerized adaptive testing could reduce the test time which is needed in classes.

In our opinion, competency assessment in music education in the 21st century is no longer possible without a technology-based setup. Without no doubt, the development and usage of these assessments is challenging, but it's worth the effort. Tbca offers new perspectives for basic and applied

research, as well as for the development of teaching concepts – and after all tbca could help to understand and facilitate students' learning in music.



Appendix – KoMus model **(translated version of Jordan et al., 2012)**

7

[7] The whole Appendix is to find on pp.72-75.

References

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The amusement park theoretical (APT) model of creativity. *Roeper review*, *27*(3), 158-163.

Barbot, B., Besançon, M., & Lubart, T. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, *52*, 178-187.

Boyle, J. D. (1992). Evaluation of Music Ability. In R. Colwell (Ed.), *Handbook of research on music teaching and learning: A project of the Music Educators National Conference* (pp. 247–265). New York: Schirmer Books.

Boyle, J.D. & Radocy, R.E. (1987). *Measurement and Evaluation of Musical Experience*. New York: Schirmer.

Colwell, R. (1999). The Future of Assessment. *Journal of Aesthetic Education, 33*(4), 53–75.

Elliot, D.J. (1995). *Music Matters.* New York: Oxford.

Harnischmacher, C. & Knigge, J. (2017): *Motivation, Musizierpraxis und Musikinteresse in der Familie als Prädiktoren der Kompetenz „Musik wahrnehmen und kontextualisieren" und des Kompetenzerlebens im Musikunterricht* [Motivation, musical practice, and the family's interest in music predict musical competence "perceiving" and "contextualizing music" and competence beliefs in music. *Beiträge empirischer Musikpädagogik, 8*, [bulletin of empirical music education research, 8], 1-21. Retrieved from: http://b-em.info/index.php/ojs/article/view/136 [15.08.2020].

Hartig, J., Klieme, E., & Leutner, D. (Eds.) (2008). *Assessment of Competencies in Educational Settings: State of the Art and Future Prospects.* Göttingen: Hogrefe & Huber.

Hasselhorn, J. (2015). *Messbarkeit musikpraktischer Kompetenzen von Schülerinnen und Schülern – Entwicklung und empirische Validierung eines Kompetenzmodells* [Measurability of practical musical competencies of students]. Münster: Waxmann.

Hasselhorn, J., & Grollmisch, S. (2014). Die Colored Music Grid (CMG) App – Ein neues Eingabeinterface zur Erfassung instrumenten-unabhängiger instrumentaler Leistungsfähigkeit [The Colored Music Grid (CMG) App – a new input interface to measure unspecific instrumental aptitude]. In W. Auhagen, C. Bullerjahn, & R. v. Georgi (Eds.), *Offenohrigkeit. Ein Postulat im Fokus* (pp. 231-234). Göttingen: Hogrefe.

Hasselhorn, J., & Lehmann, A. C. (2014). Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz (KOPRA-M) [Development of an empirically verifiable model of practical musical competence]. In B. Clausen (Ed.), *Teilhabe und Gerechtigkeit*, (pp. 77-94). Münster: Waxmann.

Hasselhorn, J., & Lehmann, A. C. (2015). Leistungsheterogenität im Musikunterricht. Eine empirische Untersuchung zu Leistungsunterschieden im Bereich der Musikpraxis in Jahrgangsstufe 9 [Heterogeneity in music class]. In J. Knigge, & A. Niessen (Eds.), *Theoretische Rahmung und Theoriebildung in der musikpädagogischen Forschung* (pp. 163–176). Münster: Waxmann.

Hasselhorn, J., & McElvany, N. (2016). Die Bedeutung außerschulischer Prädiktoren für schulrelevante musikpraktische Kompetenzen [The relevance of external pedictors for relevant-to-school musical competencies]. In N. McElvany, R. Strietholt, H.G. Holtappels, & W. Bos (Eds.), *Jahrbuch der Schulentwicklung, 19,* (pp. 168–205). Weinheim: Juventa.

Hasselhorn, J., & Wolf, A. (2017). Assessment, Bewertung und Musikkritik [Assessment, ratingand critique]. In A.C. Lehmann, & R. Kopiez (Eds.), *Handbuch Musikpsychologie.* (pp. 389-409). Bern: Hogrefe.

Hasselhorn, J., Knigge, J., & McElvany, N. (2017, October). *Förderung von Motivation und Kompetenz im Musikunterricht. Ergebnisse einer Kurzintervention* [Fostering motivation and competence in music lessons]. Lecture presented at the annual conference of the Arbeitskreis musikpädagogische Forschung (AMPF), Bad Wildbad.

Helms, D. (2005). Von Marsyas bis Küblböck: Eine kleine Geschichte und Theorie musikalischer Wettbewerbe [From Marsyas to Küblböck: a short history amd theory of musical contests]. In D. Helms, & T. Phleps (Eds.), *Keiner wird gewinnen: Populäre Musik im Wettbewerb* (pp. 11-39). Bielefeld: transcript.

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, *49*(3), 234–244.

Hickey, M. (2013). Standards, assessment, and creativity in American music education: Intersection of opportunities. In T.S. Brophy, & A. Lehmann-Wermser (Eds.), *Music Assessment Across Cultures and Continents. The Culture of Shared Practice* (pp. 15–36). Chicago: GIA.

Hornbach, C.M. & Taggart, C.C. (2005). The relationship between deve-lopmental tonal aptitude and singing achievement among kindergarten, first-, second-, and third-grade students. *Journal of Research in Music Education, 53*, 322–331.

Jank, W. (2009³). *Musikdidaktik* [Didaktique of music]. Berlin: Cornelsen.

Jordan, A.K. (2014). *Empirische Validierung eines Kompetenzmodells für das Fach Musik: Teilkompetenz „Musik wahrnehmen und kontextualisieren"* [Empirical validation of a competence model für the school subject music]. Münster: Waxmann.

Jordan, A.-K., Knigge, J. (2010). The development of competency models: An IRT-based approach to cpmeptency assessment in general music education. In.Brophy, T.S. (ed.). *The Practice o Assessmentin Music Education. Framework, Models, and Designs* (pp. 67-86). Chicago: GIA.

Jordan, A.-K.; Knigge, J.; Lehmann, A. C.; Niessen, A. & Lehmann-Wermser, A. (2012). Entwicklung und Validierung eines Kompetenzmodells im Fach Musik – Wahrnehmen und Kontextualisieren von Musik [Development and validation od a competence model in music]. *Zeitschrift für Pädagogik*, *4*, 500-521.

Keiper, S., Sandene, B.A., Persky, H.R. & Kuang, M. (2009). *The Nation's Report Card: Arts 2008 – Music & Visual Arts.* (National Assessment of Educational Progress at Grade 8). NCES 2009-488: National Center for Education Statistics.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al. (Eds.) (2003), *The Development of National Educational Standards. An Expertise*. Berlin: BMBF.

Klieme, E, & Maag Merki, K. (2008). Introduction of Educational Standards in German-Speaking Countries. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of Competencies in Educational Settings: State of the Art and Future Prospects* (pp. 305–314). Göttingen: Hogrefe & Huber.

Knigge, J. (2010): *Modellbasierte Entwicklung und Analyse von Testaufgaben zur Erfassung der Kompetenz „Musik wahrnehmen und kontextualisieren"* [Model based development and analysis of test items regarding the competence „perception of music and contextualization"]. Bremen: Universität Bremen. Retrieved from: http://nbn-resolving.de/urn:nbn:de:gbv:46-diss000120066 [15.05.2020].

Knigge, J. (2014): Der Kompetenzbegriff in der Musikpädagogik: Verwendung, Kritik, Perspektiven [The term „competence“ in music education]. In Vogt, J.; Brenk, M. & Heß, F. (Ed.), *(Grund-)Begriffe musikpädagogischen Nachdenkens – Entstehung, Bedeutung, Gebrauch* (pp. 105-135). Münster: LIT.

Knörzer, L., Stark, R., Park, B., & Rolle, C. (2016): “I like reggae and Bob Marley is already dead”: An empirical study on music-related argumentation. *Psychology of Music*, *44*(5), 1158-1174. DOI: 10.1177/0305735615614095

Kormann, A. (2005). Musiktests [Music tests]. In. Deutsche Gesellschaft für Psychologie (Ed.), *Enzyklopädie der Psychologie: Vol. Serie VII, Bd. 2. Spezielle Musikpsychologie* (pp. 369–407). Göttingen: Hogrefe.

Lehmann, A.C., & Ericsson, K.A. (1997). Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students. *Psychomusicology*, *16*, 40–58.

Lothwesen, K., & Lehmann, A. C. (2017). Komposition und Improvisation [Composition and improvisation]. In A.C. Lehmann, & R. Kopiez (Eds.), *Handbuch Musikpsychologie* (pp. 341-366). Bern: Hogrefe.

McPherson, G.E. (1995). The Assessment of Musical Performance: Development and Validation of Five New Measures. *Psychology of Music, 23,* 142–161.

Niessen, A., Lehmann-Wermser, A., Knigge, J., & Lehmann, A. C. (2008). *Entwurf eines Kompetenzmodells 'Musik wahrnehmen und kontextualisieren'* [Proposal for a competence model of „music listening and contextualizing]. *Zeitschrift für Kritische Musikpädagogik, Sonderedition: und Kompetenzmodelle für das Fach Musik?*, 3–33. Retrieved from: http://www.zfkm.org/sonder08-niessenetal.pdf. [15.08.2015].

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* New York: Guilford Publishing.

Venus, D. (1969). *Unterweisung im Musikhören* [Instruction for listening to music]. Wuppertal: Henn.

Weinert, F.E. (2001). Concept of Competence: A Conceptual Clarification. In D.S. Rychen, & L.H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Göttingen: Hogrefe.

West, M. L. (1992). *Ancient Greek music*. Oxford: Clarendon press.

Wilson, M. (2003). On Choosing a Model for Measuring. *Methods of Psychological Research Online, 8*(3), 1–22.

Wu, M. L., Adams, R. J., Wilson, M. R. (1998). ConQuest: Generalised Item Response Modelling Software. Melbourne, Australia: Australian Council for Educational Research.

Appendix

	D1: Perception and musical memory
680 670 660 650 640 630 620 610 600 590 580 570 560	**Level III:** Students have the competency to perceive music which has a high level of complexity. E. g., they can follow several voices simultaneously and identify starkly altered motivic material in new contexts. They are able to grasp music-theoretical phenomena such as major and minor mode.
550 540 530 520 510 500 490 480 470 460 450 440 430	**Level II:** Students can also identify and isolate musical phenomena in more complex contexts. They may recognize single instruments in an ensemble sound and identify rhythmic groupings or simple melodies in semi-complex music and its variations. In addition, they can describe the relationship of musical phenomena (e.g. A-A-B-A form, comparison of pitches). This competence hints at a heightened performance of musical memory, since information has to be retained over longer time spans and/or despite interferences. Furthermore, students are also able to identify the musical expression of a piece of music by relating the piece's musical parameters to the perceived expression.
420 410 400 390 380 370 360 350 340	**Level I:** Students recognize and distinguish sallent features in listening (e.g. clear dynamic gradations, different sounds and instrumentation, respectively). They can also identify sections of simple and unambiguously segmented formal structures. Moreover, students grasp prototypical musical expression (e.g. festive or sad). At this level students only engage in elemental performance of musical memory by identifying musical phenomena after a short delay.

Appendix – KoMus model **(translated version of Jordan et al., 2012)**

	D.2: Terminology
680 670 660 650 640 630 620 610 600 590 580 570	**Level III:** Students possess a well-developed ability to verbalize. They can not only assign given descriptions to corresponding perceptions, but, in addition, describe music adequately and distinct in their own words. For this, they employ appropriate jargon. This ability combines with a developed competency of perception. Thus, students may be able to isolate and describe separate musical parameters from an integrated sound. They are also able to critically evaluate a musical performance using adequate terminology and provide appropriate feedback to the performer.
560 550 540 530 520 510 500 490 480 470 460 450 440 430	**Level II:** Applying basic music theoretical jargon intentionally (e.g. forte/piano, canon) in response to perceived events is characteristic for the competence of students. They identify and name even less familiar instruments (e.g. the clarinet) and are able to superficially describe musical performances using mainly everyday language.
420 410 400 390 380 370 360 350 340	**Level I:** Students can recognize salient music features, like clear dynamic gradations and are able to describe them with the help of everyday words. They can identify well-known instruments through listening and assign the correct names to these instruments (e.g. piano, guitar).

Appendix – KoMus model **(translated version of Jordan et al., 2012)**

	D3: Notation
680 670 660 650 640 630 620 610 600 590 580 570 560 550 540 530 520 510 500	**Level II:** Students can follow complex traditional notation, for example score notation, with regard to meter as well as pitch contours. For this, they rely on music-theoretical knowledge (e.g. for indications). This compentence combines with an increased ability to perceive and remember music, i.e. students also track challenging musical events and sequences in notated form. In addition, they are able to recognize more complex rhythmic phenomena in traditional notation such as syncopations and write simple rhythms themselves.
490 480 470 460 450 440 430 420 410 400 390 380 370 360 350 340	**Level I:** Students can match their perception of sallent musical features to graphical and simple traditional notation. They possess a basic understanding of the traditional concept of notation (with notated time sequence in horizontal and pitch sequence in vertical arrangement), allowing them to orient themselves in music notation using melodic contour. Students are thus able to assign a sounding melody to its correct notation, they are able to follow its presentation, and they can identify locations in a familiar melody where false notes have been introduced.

Appendix – KoMus model **(translated version of Jordan et al., 2012)**

	D4: Historical/Contextual knowledge
680 670 660 650 640 630 620 610 600 590 580 570	**Level II:** Students have extended knowledge of music styles and genres as well as terminology, what makes it possible to describe the perceived music (e.g. jazz or rock) differentiated with regard to their stylistic characteristica. Moreover, they are able to order musical examples in a historically correct manner and assign them to periods and cultural contexts if they are clearly distinguishable and characteristic for their date of origin. Moreover, students have the competence to describe without help the stylistic, cultural and historical specifics of music. Finally, they can consider functional aspects of music, such as the function of music as background.
560 550 540 530 520 510 500 490 480 470 460 450 440 430 420 410 400 390 380 370 360 350 340	**Level I:** Students have the competence to connect their music perception with cursory, partial contextual knowledge about music history, musical styles, as well as everyday social and cultural musical contexts. Hence, they succeed in assigning musical examples to popular musical periods and everyday usages. For this, they activate a rather vague and implicit understanding of important concepts (therefore graphical representations seem to support problem solving).

Appendix – KoMus model **(translated version of Jordan et al., 2012)**

Marina Gall

5. Assessment Tools and Practices in England: What Have We Learned?

In this chapter, writing as an ex-classroom music teacher and current teacher educator[1] and researcher, I reflect upon assessment in music in the context of music education in England[2]. As will be explained later, the main focus will be on group composing in secondary school, for students aged 11 to 14 years of age, which will also include a consideration of assessment tools and practices when technology is used in the music classroom. The chapter begins by providing information on classroom music in England and explaining the position of music technology within this. I then present an historical context to current assessment practices. After considering why we assess in music, I reflect upon current assessment practice in England, and then present ways in which technology impacts on this. The chapter ends with a short conclusion in which I make suggestions of further readings on the subject of assessment of musical learning in schools.

1 I am aware that the terms 'teacher trainer' and 'teacher training' are used in many European countries; I will use the terms commonly used in England: 'teacher educator' and 'teacher education'.

2 Education systems are different in each country across the UK: England, Northern Ireland, Scotland and Wales.

Music Education in Schools in England

Stages of education in England (and in Wales and Northern Ireland) are called 'key stages' (known as KS) and are organised as follows:

	Period	Age of Children	Is Music Compulsory?	School Music Level Examinations
Primary School				
Key Stage 1	Years 1-3	5–7	Yes	None
Key Stage 2	Years 4-6	7–11	Yes	None
Secondary School				
Key Stage 3	Years 7-9	11–14	Yes	None, although until 2014 schools reported to the government at the end of KS3 (see later discussion)
Key Stage 4	Years 10-11	14–16	No	Typically, General Certificate of Education (GCSE) or Business and Technology Education Council (BTEC)[3]
Key Stage 5	Years 12-13	17-18	No	Typically, General Certificate of Education (GCE – commonly known as Advanced or A level) or BTEC

Table 1: Organisation of the 'key stages'

3 BTEC courses and usually more vocational than those of GCSE (Pearson 2019b) and A level (Pearson, 2019a), often including a considerable amount of project work, as opposed to the mainly examination-based GCSEs and A levels. See examples at Pearson, 2019c.

Music in English primary schools at key stages 1 and 2 is organized differently to that in secondary schools at key stage 3, reflecting the distinct organizational structures, differences in available resources and the diversity of teachers' experiences of music in those institutions (Gall & Breeze, 2007). Opportunities for music in the primary school are variable (Henley & DCMS, 2011); constraining factors are teachers' lack of music expertise and/or confidence, arising from the, typically, limited teacher training time for music provide for primary teachers (Ballantyne & Packer, 2004; Garvis, 2013) and a lack of time allocated to the subject in many schools (Adams, 2019). As such, some students can leave primary school with very little experience of music-making and few skills.

Secondary schools usually employ at least one music specialist who teaches music throughout the school. At key stage 3, about an hour a week for Music used to be the norm (DfES, 2002), although, more recently, this has decreased in many schools (Norris, 2018).[4] More hours are allotted for students studying for examination courses after the age of 14.

The first National Curriculum for Music (for students aged 4-14) in England was presented in final form in 1992 (Finney, 2016). The latest version is diminished in detail in comparison with the previous four editions (DfE, 2013b & 2013c; also presented in Appendix 1). It provides a very broad curricular framework for all pupils aged 5 to 14 in state-maintained schools. At primary and at secondary schools, until the age of 14, students are required to develop the skills of performing, composing, and listening and, through these, to advance their appraisal skills and also their ability to apply knowledge and understanding (Finney, 2016). Since the advent of the earliest National Curriculum for Music, the *interrelated* nature of composing, performing and listening has been, and still is, regarded as fundamental to pedagogy, requiring

4 See reasons in Burns (2017) The Guardian (2018). Furthermore, I am aware of schools in my locality in which music forms part of a 'carousel' with drama and art: students study each for only one term each academic year, at age 11-14.

the need for an integrated approach to skills development (Adams, 2001; Fautley & Savage, 2008; Finney, 2011). Accordingly, since the late 1980s, there has been a strong focus upon practical music making within classroom music, with contextual historical information and theory introduced only to deepen musical understanding and skills (Spruce, 2016).

A typical lesson for students aged 11 to 14, focuses on composing or performing tasks, with students either 1) working in groups of three to six, using music keyboards and/or voices and/or acoustic (including Orff) instruments or two) composing and performing in pairs, using computer workstations. Although iPads are available[5], my work in local schools indicates that it is rare for a music teacher to have access to these, hence the continued use of computer workstations. World music, jazz and pop are now commonplace within the curriculum (DES, 1991). With no detailed government prescription of content, teachers are free to create lessons of their choice; many design their own schemes of work and materials (Spruce, 2016). In the last decade, there has also been a strong focus upon student autonomy in relation to pedagogical approaches (D'Amour, 2009; Harrison & Finney, 2010). Statutory requirements for Music exist for children up to the age of 14 (the end of key stage 3). After this, students make choices about the subjects they wish to study. If they select Music at key stages 4 and 5, students work towards national public examinations as outlined in Table 1 above. These place the same focus upon active music making that one finds within the National Curriculum for younger pupils. However, government examinations require that all student work is carried out individually. As such, assessment processes for composing for 11 to 14-year-olds have had to be conceived differently to those of the formal GCSE and A level school examinations.

[5] In many secondary schools there are sets of iPads, held centrally, available through a booking system. However, these are often difficult to acquire by music teachers because they are needed on a regular basis, and over an extended period of time.

Technology in the Classroom

Music technologies first came into English schools as a result of developments in public examination courses for 17 to 18-year-old students. In the early 1990s only one Advanced level course, entitled *Music* was available for students aged 16-18. In this, there was a strong focus upon traditional harmony and aural skills, analysis, performance and classical music. Later, an optional composition element was introduced, and became mandatory in the 1990s (Paynter, 1992). This element became the focus of use for computers. Furthermore, in 1995, a completely new type of A level examination - *Music Technology* - was offered (Kwami, 2001). This differed considerably from what is known as the *Music A* level since it was designed to develop students' knowledge and use of computers and recording hardware and software. As well as a composing task, the *Music Technology* syllabus of the time included sequencing, multi-track recording and questions on the principles of music technology, some related to aural extracts. The current syllabus, essentially, remains the same. In recent years, additional BTEC courses, with strong vocational foci, have also evolved (e.g. Pearson, 2019).

At first, as a result of changes to A levels, in Music lessons, computers were almost exclusively used by older students. However, after 1985 when composition became a central aspect of the GCSE (Atkinson & Spruce, 2002), in schools that owned sufficient resources, some 14 to 16-year-olds also used music sequencing for creative purposes. Later modifications to the GCSE examination syllabi saw sequencers, multi-track recorders, samplers and record decks being permitted for not only for composing purposes, but also solo and ensemble performing (Edexcel, 2000, p. 15).

Up to, and for a considerable time into, the new millennium, computer resources were limited in most schools (Ofsted, 2004) despite a requirement, in the revised 1999 National Curriculum for Music, for students to use 'ICT to create, manipulate and refine sounds' (DfE/QCA, 1999, p. 30). Whilst there still remain difficulties in working with technology in classes for 11 to

14-year-olds (Gall, 2013), over time, more Music departments have become equipped to enable whole class work with groups of up to 30 students. At present, throughout key stages 3, 4 and 5, music technologies are mainly used for composing, although they are used by older students for a range of purposes related to their Music examination work.

Assessment in Music Until 2014

To fully understand assessment in Music in 2019, one must be aware of the governmental assessment system that prevailed until 2014. This took the form of target levels, specific to each school subject, being designed for use by class teachers (DfE/QCA, 1999; also provided in full in Appendix 2). These target levels were originally intended to be used to provide information to the government, on each individual 14-year-old, at the end of KS3 (Fautley and Savage, 2011). Those for Music were expected to be reported on holistically, without a separate level being defined for each aspect: composing, performing and listening (Fautley, 2010). However, in order to gratify the data-driven climate of education in England in which assessment fulfilled a systemic need for whole-school auditing, rather than for learning purposes (Mansell, 2007; McClean, 2016), across many subjects, teachers misused the system, subdividing the levels[6] in order to demonstrate student progress over short periods of time. Indeed, in 2011, Fautley and Savage reported on the substantial use of sub-levels for grading a *single* piece of work, noting that only 9% (of 104 teachers) were using the level system in the way it was originally intended. This was despite the government highlighting the need for consideration of integrated practice and stating that teachers should avoid using levels to assess isolated activities (Ofsted, 2009). Considering the levels in Appendix 2, the reader may well appreciate the difficulty in their use, even without any sub-levelling: how was a teacher expected to arrive at

[6] This was often into a further 3 categories, although some schools created even more (Fautley and Savage, 2011). See Ofsted, 2012a, p.37 for an example of smaller sub-divisions.

one grade for attainment when such differing aspects of music were being considered? I remember, in years past, being faced with an impossible decision to make when allocating a level to a student who was a superb performer on the cello, but who, interestingly, did not have strong aural skills and was weak at composing. Was I supposed to aggregate marks related to each of the three skills which would lead to his being presented as an 'average' student (level 5/6) or state that he was 'exceptional' (beyond level 8) and misrepresent his aptitude in two of the three key areas being assessed?

A further issue was the expectation in many schools that a child makes linear progress in Music (Cain, 2001; Fautley, 2017), a notion contrary to the thoughts of music educationists (Adams, 2001; Cain, 2001; Mills, 2005; Finney, 2011; McClean, 2016). Additionally, this progress was always expected to be upwards (Fautley, 2010). One wonders how this can be when students' preferences for different styles of music are such that they might well attain highly in a project of their interest, compared to one carried out at a later stage which excites them less.

Assessment in Music from 2014

In 2013, with reforms to the National Curriculum across all subjects, the government announced the end to statutory assessments using the levels and handed the responsibility of what, when and how to assess to schools (DfE, 2013a). As can be imagined, this was a huge change which, in reality, needed considerable time to carry out effectively, yet many schools, quickly created replacements similar to the previous level-based system (Christodoulou, 2015). Additionally, despite concern by the government (Ofsted 2012b), many of these newly-created approaches were 'whole-school', led by school managers whose designs did not align well with music subject-specific processes and tools (Balcombe, 2016). This led to many music teachers adopting approaches they knew were not appropriate, but which were expected by their school leaders, whose aim was homogeneity (Gall, 2015). Given that

the last classroom inspections by specialist music educators took place in 2011, no governmental support was available to Music teachers to argue their case for assessment systems suited to their subject. Furthermore, it is interesting to ponder whether non-specialist school inspectors had sufficient knowledge to make judgements on music teachers' competence generally and in relation to student progress which, at this time, was a key focus. Ironically, a government report had itself highlighted the dangers of non-specialists commenting on music teachers' capabilities:

> *Too many [non-specialist school] senior staff were not well enough informed about the key features of effective learning in music to make accurate judgements about the quality of teaching. Too often, their observations gave more importance to generic teaching strategies than to the musical qualities of teaching and learning.* (Ofsted, 2012a, p. 57)

Why Assess?

I have included this lengthy explanation of assessment in England over the past decades, in order that music educators in other countries might learn lessons from our mistakes.[7] For years, accountability has driven the reporting of progress, with schools impelled by expectations of what government inspectors want to see (McIntosh, 2015). Clearly, there will always be a need to report to the government and to parents, thus teachers will be required to record attainment. However, surely, the crucial reasons for assessing are to provide feedback to students so that they are clear about their current attainment and how to progress further; and to inform the teacher's planning of subsequent learning. Despite the many difficulties brought about by national

[7] These were even acknowledged by the government: reporting on the decommissioning of the level system, the, then, Secretary of State described it as 'disjointed' (NAHT, 2014, p.6) also '… complicated and difficult to understand... It also encourages teachers to focus on a pupil's [sic.] current level, rather than consider more broadly what the pupil can actually do' (DfE, 2013a).

and school-based initiatives, highly effective strategies for assessing musical progress and attainment have been adopted in many schools. Central to these is formative assessment.

Formative Assessment

In this chapter, 'formative' assessment is used to denote 'taking learning forward' (Fautley & Colwell, 2018) and revolves around the key principles of 'Assessment for Learning' (known as AfL) suggested by the Assessment and Reform Group (ARG, 1999):

- the provision of effective feedback to pupils
- the active involvement of pupils in their own learning
- adjusting teaching to take account of the results of assessment
- a recognition of the profound influence assessment has on the motivation and self-esteem of pupils, both of which are crucial influences on learning
- the need for pupils to be able to assess themselves and understand how to improve.

What do these mean in English Music classrooms? I now consider formative assessment within composing lessons for 11 to 14-year-olds.

In England, the most common approach to supporting students in all practical activities, including composing, is for the teacher to move between the student groups as they work, offering advice when necessary. At best this does *not* take the form of the teacher leading or providing their own ideas. It involves, first, observing the young people as they work, and then engaging in dialogue with the pupils to gain an insight into *their* thoughts about the music and the next stages they perceive to be necessary. Indeed, careful decision-making about when to intervene whilst the students are working is key to the creative process. This practice of helping students to critique their own work and *drawing out* their ideas, rather than the teacher providing a way forward,

has been recognized as important by many (Black et al., 2002; DfES, 2004; Odena, 2012; Ofsted 2012a) including the government who, following their report on visits to schools between 2008 and 2011, suggested that:

> *'The most effective assessment practice observed helped students to listen more accurately to their own work, helped them identify for themselves where improvements were needed…'.* (Ofsted 2012a, p.38)[8]

Another key element within today's key stage 3 Music lessons is peer assessment, which is also found to stimulate students' attentiveness to their own progress (Black et al. 2002). The inclusion of 'work-in-progress' sessions is commonly used as a means of helping students to fully understand expectations: a group performs their composition to the rest of the class who provide oral feedback, mediated by the teacher. As I observe my trainee teachers in school classrooms, I am constantly reminded of how motivating it is for the young people to see and hear their classmates' work, and for students' own music to act as 'models' for their peers. Aside from the benefits to composing, this practice also offers informal opportunities for students to develop listening and performance skills. Furthermore, as Kordes et.al. note, peer feedback helps the development of 'learning skills, critical thinking, evaluation skills…' (2014, p. 296). With good teacher facilitation, through such activities, students are also learning a key life skill of how to present opposing ideas to others, in a respectful way. In most classrooms I visit, staff teach the students to first offer positive points about the work and then to suggest some targets for improvement. Common approaches are for students to state 1) 'what went well' (positives) and 'even better if' (things to work on further) or 2) '3 stars' (3 good points about the work) 'and a wish' (a wish that an aspect might be better).

8 The discussion demonstrates why Swanwick suggested '…to teach is to assess' (1988, p.149).

The above discussion has identified how students can be actively involved in their own learning. However, this can only come about when there is a clear understanding between the teachers and the young people of exactly what learning is expected within a given project. A common approach has been criterion-referenced assessment, which Kempa and L'Odiaga explain as 'the measurement of students' performances against a set of criteria specifying educational attainments and ability levels (1984, p. 56). Criteria created for assessment purposes are expected to denote not only *what* is to be produced but also the *quality* of the outcome. In England, it has been customary to present three possible outcomes to the students, indicative of different levels of attainment in relation to that specific product.[9] Within a scheme of work in which students are to compose music to fit episodes from the film Romeo and Juliet, these might be:

- has created music which largely reflects actions in the film and includes the use of the original character motifs; the music is mainly based around sound effects;
- has created music which largely reflects the atmosphere in the film, and includes appropriate modifications to the character motifs;
- makes subtle use of timbre, texture, rhythm and harmonies to create music which reflects the atmosphere in the film; this includes creative modifications to the character motifs.

However, there is considerable criticism of criterion referencing in relation to composing. Spruce suggests that 'rigidly criteria-related' assessment models can lead to the risk of a fragmentation of the compositional process (1996, p. 175). Thorpe (2012) adds that prescriptive assessment could inhibit learning because pupils become overly concerned with accomplishing the success criteria, as opposed to being creative (ibid.). Salaman (2008) agrees,

9 These are conceived as follows: outcome b) relates to what most students in the class might be expected to attain, outcome a) to what might be expected of some students who have more difficulties, and outcome c) to the work of certain higher attaining students.

suggesting that, by presenting criteria, the music teacher is 'expecting the composer to conform to preconceived yardsticks of compositional virtue before the work has begun' (ibid., p. 239). Others join him in their concern that criterion-referencing can marginalize learning that takes place outside of the pre-ordained criteria: students might well create high quality work without fulfilling the proposed criteria (Fautley, 2010; Balcombe, 2016). The above film music example could be a case in point: a highly appropriate musical soundtrack could be created without including character motifs.

A more recent approach to criterion-referencing has been posited by Fautley and Daubney; they suggest creating a criterion, then measuring against this on the following three-point scale: 'working towards' (is not yet able to do it), 'working at' (can do it) and 'working beyond' (can do this competently) (2015, p.12). Within each project, they advise creating four or five criterion statements that are specific to that topic (ibid.). Interestingly, this three-point approach is a return to what was expected in the earliest days of the National Curriculum (Pratt & Stephens, 1995).

How to Record and Use Formative Assessment Information

So how does one capture learning that takes place in lessons? There will always be a requirement for some form of record of individual progress and attainment, but what type is useful, and also practical, for staff who may teach a class of up to 30 students, for only 50 minutes, once a week – and have 10 such classes on their timetable? In recent years, I have seen Fautley and Daubney's radar graph charts (2015, pp. 17-19) – often modified in form – used very successfully by a range of experienced and novice teachers.

Below is an example, from a local school, related to a song writing project in which students, aged 13-14, worked in pairs. As is clear, aspects of the work are each graded out of a possible five marks (rather than the three suggested above). Assessment criteria, related to each mark, were provided to help the young people self-assess, over time, as they composed their song.

Over the weeks, as students either added more to their song, or raised the quality of one aspect, they noted this in blue (self assessment).[10] The Head of Music who designed this approach also experimented with the addition of a 'notes' box to one side of the sheet, intended for the purpose of self-assessment: She encouraged the students to write 1) comments on ideas they explored, including any that were rejected and 2) their thoughts on what was positive about the music and what needed further consideration. She also used this box to offer teacher feedback at the end of the project.

As is common in England when assessing 11-14 year olds' practical music tasks, in this project the students presented their final work to the whole class, and the class together made judgements on the final grades (peer assessment). These, ratified by the teacher, are the marks, in red in Figure 1 (summative assessment used for reporting purposes).[11]

[10] On the student document there are further 'spoke' lines from the centre to the outside of the diagram making it easy for them to plot their marks on the diagram.

[11] The mark for team work and listening and appraising were decided on through discussion between the teacher and the pairs of students

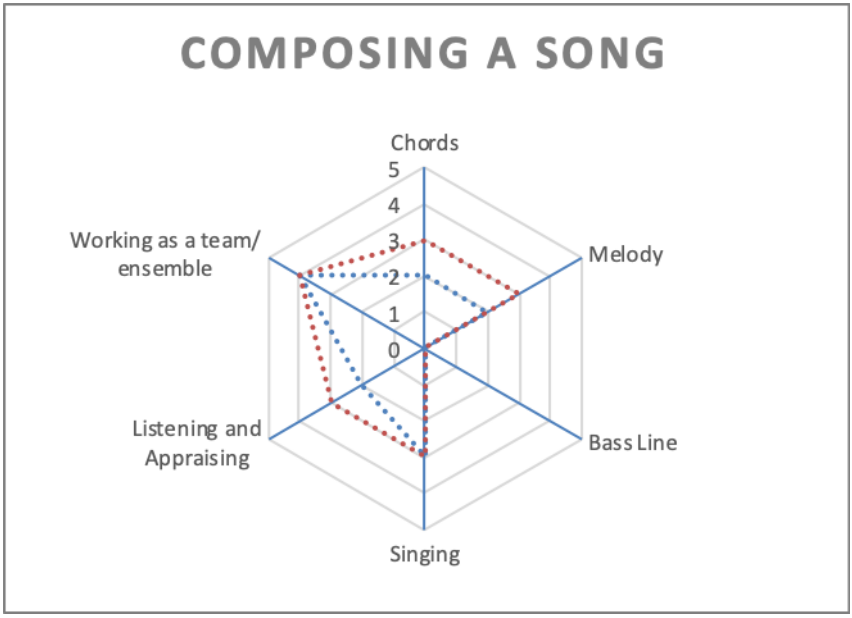


Figure 1. Example of a student radar graph assessment form

This system was used for all projects at key stage 3. Students claimed that this visual representation helped them to understand what was expected of them in terms of the product to be created and provided a self-assessment system that was clear to follow (McLean, 2016).

As discussed earlier, a downside to the approach is that it is atomistic: a final composition could be creative and high quality even if it did not include chords and a bassline. However, there are many positive aspects to its use. Firstly, group composing activities necessitate the teacher moving around, focussing on one set of students at a time. As such, it is not always easy for her/him to be clear about what has been created by whom in other groups. Placing the onus on the student to complete the form greatly benefits the teacher by providing her/ him with information that otherwise might have

been missed.[12] In this project, since the students were given autonomy in deciding how to approach their work – that is, the order in which they devised the parts of the song – and were encouraged to self-assess as the work progressed, the form also provided the teacher with some information about the composition process of each pair.[13] Moreover, the version of the radar assessment form including a written 'notes' section, offers the potential for further insight into how the final piece came about. Breeze explains the importance of this:

> *…the experimental phase involves many musical transformations that do not necessarily appear in the final composition product; therefore, if the product alone is considered for assessment, the teacher is not fully informed about the learning potential of the group, which will in turn affect future planning.* (Breeze, 2011, p.403)

Another method of noting information about the process of composing, which I encourage my trainee teachers to use whilst on teaching practice, is presented in Appendix 3 (an example, developed and used by one trainee teacher, can be found in Appendix 4)[14]. A sheet is created for each class and the novice teacher uses it to make notes on individual students' progress and attainment. The idea is that it is kept on the teacher's person, for use whilst the class are engaged in practical work and provides supplementary information to any form (such as the radar assessment form) that the students themselves are using. Trainee teachers do not have to use it in every single lesson (although the person who created the example in Appendix 4 did), but many find it helpful in gaining information, in the 'here and now', about

[12] It is interesting, and pleasing, to see how honest students are in registering their personal contributions to group work.

[13] Some students were noted to begin by creating the lyrics and melody then added chords; others began with a chord sequence, then developed their lyrics and melody to fit this.

[14] This sheet relates to a composing project using GarageBand. The detail arose from there being no other method of collecting formative assessment information in the trainees' school.

difficulties that certain students are facing or on pupils who may be nearing completion of the set tasks. This detail concerning specific students can then be used to inform future planning. A further attribute is that the form provides easily accessible detail on pupils' current musical experience and skills, and/or additional learning needs, which can be very useful when interacting with groups and individuals. Whilst, here, I have explained the significance of this document to the teacher, it should be noted that, the majority of the time, the information placed on the sheet arises not only from observation but also from discussions between the teacher and the groups of young people, during their engagement in the practical tasks. These will, naturally, provoke consideration of the merits of the music, and ways in which the students can further develop their pieces.

Progression

Having discussed the gathering of formative assessment information, I now move to a consideration of musical progression: 'the cumulation of attainments over time, evidenced through musical endeavour' (Fautley and Daubney, 2015, p. 17). In Fautley and Daubney's document '*The National Curriculum for Music: A Framework for Curriculum, Pedagogy and Assessment in Key Stage 3 Music'* (ibid.) one can see how composing can be considered alongside other key musical skills so as to consider learning holistically. It also explains how data from ongoing musical assessments involving the range of musical skills can be used to evidence progress, and suggests that the radar graph system, slightly differently conceptualized to the example above, can be used to capture this progress (ibid., also see Figure 2).

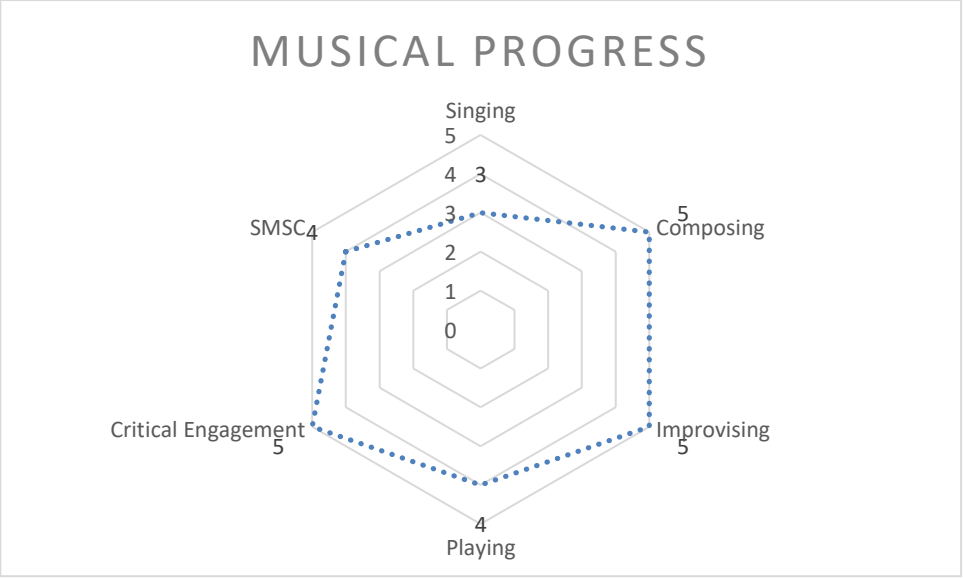


Figure 2: Modified version of Figure 4 from Fautley and Daubney (2015, p.19)[15]

A further key feature is that this presentation of an individual's musical learning can be of use both summatively and formatively: it can be used at a given time to provide data for whole school auditing or reporting purposes and is also a starting point for individual pupils to reflect upon possible personal targets within the next musical activity. Once again, this requires a consideration not only on which skill(s) to target but also a clear understanding, by the student, of how to go about developing those skills.

Technology In and For Assessment

It is perhaps self-evident that many of the assessment processes and practices previously discussed are as applicable to music activities which include technology as to those that do not. However, the emergence of a wide range of

[15] SMSC means spiritual, moral, social and cultural development (See SMSC, undated).

generic and music-specific technologies do provide some differing affordances. In the following reflection, I consider their use in two differing contexts.

Technology In Assessment

As explained previously, in England, technology is used by 11 to 14-year-olds mainly as a 'tool' for composing. I use the term 'tool' since, as should be self-evident, technology skills are not of import in their own right but can be hugely beneficial to a composer. In school, when a teacher assesses a composition, s/he judges it on the merits of the composition itself not in relation to the component technological aspects. However, developing technological skills has been seen to be key to the engagement, in composition activities, of a wide range of students, including those who might otherwise lack interest or confidence (Mills & Murray, 2000; Gall & Breeze, 2007; Wise et al. 2011). As such, a number of teachers with whom I work see an audit of specific technology skills as essential. Beginning this at the start of a student's time in secondary school is helpful in gaining baseline information. Given the range of experiences and resources that the young people may have in their lives outside of school, this then enables the teacher to plan work for their class with an awareness of individual competence. Furthermore, an audit with individual sheets for students to map their own progress offers the same benefits as the radar forms described earlier: logging their development of skills motivates the children and provides important data for the teacher. Appendix 5 provides a current example from a local school, related to one unit of work in which pupils composed with sequencing software on MAC computers.[16]

16 The addition of a box to indicate the date when the student acquired the skill would offer an even more detailed picture of student capabilities over time.

Technology For Assessment

Technologies are helpful in aiding teachers both in formative and summative assessment. Many are commonplace in today's English schools. For years, musical recordings of compositions have been made, using varied equipment. The value of these, for summative assessment purposes, is recognized. As one governmental report notes:

> *A well-ordered catalogue of recordings over time, supported by commentaries and scores, provides a very effective and compelling way to demonstrate students' musical progress.* (Ofsted, 2012a, p.38)

Furthermore, technologies have brought about hugely significant changes within formative assessment. In my early research on composing on computers in the classroom, my colleague and I noted the significance of the students being able to save their work on networked PCs: pupils were able to access and, therefore, peer assess the work of all their classmates - without moving from their own seats (Gall & Breeze, 2005). In such a context, the teacher can either run peer assessment as a whole class venture or as an activity only for certain pupils, thereby having more flexibility in the use of classroom time. Furthermore, music created in each lesson, saved by the students as discrete files clearly dated, can be used by the teacher to consider detail of a student's process of composing – including musical material that was rejected.[17]

For students composing in groups with acoustic instruments and voices, the ubiquity of mobile phones is helpful to formative assessment. Students are now able to record their work at any point in the composition process, whether or not a member of staff is present. Whilst engaged in group work, students cannot easily get a sense of the piece of music as an entity. Recording and listening back to interim work – whether in audio or audio-visual

[17] Earlier versions of music are also helpful when pupils are absent for the final lesson of a composing project.

form – offers the potential for students to gain a stronger notion of the composition as *a whole* and to realize, for themselves, the changes / developments that are necessary.Video footage is also of particular import to teachers using the tool for summative assessment of composing since individuals' contributions are, evidently, clearer when visual data is available as well as the musical product.[18]

Thus far we have focussed on students and teachers capturing and sharing work within the classroom. Latest developments in collaborative technologies have resulted in the availability of online platforms that also enable access outside of school (Edmodo, 2019; Firefly, u.d.). School colleagues with whom I work have discussed the many merits – to students' learning and to assessment – through the use of Edmodo (Gall, 2015). Students' practical work is regularly uploaded to this platform for easy storage and access, inside and outside school. One teacher said that his music homework is always to request that the students go to the platform and peer evaluate at least one other group's music. He finds that students really enjoy this, and often produce more work than he had expected. A cycle is produced by including this pupil feedback at the start of the next lesson, enabling the teacher to highlight comments that are appropriate and to further explore or rectify any misconceptions. The fact that the young people's parents are also permitted to access the site means that they can see their child's work and progress without having to wait for the annual visit to school, following the school's formal report (Gall, 2015).

Final Comments

In this chapter we have seen that, over the last two decades in England, music educators, of necessity, have had to place a sharp focus on assessment. A lot has been learned over the period. Martin Fautley and Ally Daubney, in

[18] This is even more significant in group performing activities in which gesture, eye contact and gaze are important.

particular, have been of huge significance to music teachers, teacher educators and researchers in their thinking and practice. Their latest publications deserve close attention since they provide detailed frameworks for assessment in music for 11 to 14-year-olds (Fautley and Daubney, 2015) and also in the primary sector (Daubney, 2017). As final thoughts, I suggest that those wishing to make changes to their own school assessment system take heed of a concern expressed by Black et al. in 2001:

> *…we start out with the aim of making the important measurable and end up making only the measurable important*. (Black et. al., 2001, p.58)

A focus on musical experiences and not 'the mechanics of assessment procedures' (Ofsted, 2012a, p. 37) is a basis for classroom music lessons in which students engage, enjoy themselves and make progress.

References

Adams, P. (2001). Assessment in the music classroom. In C. Philpott (Ed.) *Learning to Teach Music in the Secondary School: A companion to School Experience* (pp. 163-176). London: RoutledgeFalmer.

Adams, R. (2019*). 'We fight for every penny': teachers say not enough resources for arts education*. *The Guardian*, 11 January 2019. Retrieved from: https://www.theguardian.com/education/yougov-polling-blog/2019/jan/11/we-fight-for-every-penny-teachers-say-not-enough-resources-for-arts-education [21.01.2019].

Assessment Reform Group (ARG) (1999). *Assessment for learning – Beyond the black box. Cambridge: University of Cambridge School of Education.* Retrieved from: http://www.nuffieldfoundation.org/assessment-reform-group [21.01.2019].

Atkinson, P. & Spruce, G. (2002). Music Education & GCSE Music. In G. Spruce (Ed.), *Aspects of Teaching Secondary Music: Perspectives on Practice* (pp. 216-227). London: RoutledgeFalmer.

Balcombe, O. (2016). *What effect has the removal of the statutory assessment system of National Curriculum Levels had on music education at Key Stage 3: How is progression being mapped? A case study.* Unpublished Master's Thesis. University of Bristol.

Ballantyne, J. & Packer J. (2004). Effectiveness of preservice music teacher education programs: Perceptions of early-career music teachers. *Music Education Research*, *6*(3), 299-312.

Black, P., William, D., Harrison, C. & Marshall, B. (2002). *Working inside the Black Box: Assessment for Learning in the Classroom*. London: King's College London.

Breeze, N. (2011). Multimodality: an illuminating approach to unravelling the complexities of composing with ICT?. *Music Education Research, 13*(4), 389-405.

Burns, J. (2017). *Music could face extinction' in secondary schools. BBC News,* 9 March 2017. Retrieved from: http://www.bbc.co.uk/news/education-39154242 [21.01.2019].

Cain, T. (2001). Continuity and progression in music education. In C. Philpott & C. Plummeridge (Eds.), *Issues in music teaching* (pp. 105-117). London: RoutledgeFalmer.

Christodoulou, D. (2015). Assessment Knowledge, in J. Simons & N.Porter (Eds.) *Knowledge and the Curriculum* (pp. 21-28). London: Policy Exchange.

D'Amour, A. (Ed.) (2009). *Musical Futures: An approach to teaching and learning. Resource Pack.* (2nd ed.). London: Paul Hamlyn Foundation. Retrieved from: https://www.musicalfutures.org/resources [21.01.2019].

Daubney, A. (2017). *Teaching Primary Music*. California / London: SAGE.

Department for Education (DfE) (2004). *Pedagogy and Practice: Teaching and Learning in Secondary Schools. Unit 12: Assessment for learning.* Retrieved from: https://webarchive.nationalarchives.gov.uk/20110809174429/http://wsassets.s3.amazonaws.com/ws/nso/pdf/2deff878cffd2cdcd59a61df29e73105.pdf [21 January 2019].

Department for Education (DfE) (2013a). *Assessing without Levels*. Retrieved from: http://webarchive.nationalarchives.gov.uk/20130123124929/http://www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00225864/assessing-without-levels [21.01.2019].

Department for Education (DfE) (2013b). *Music programmes of study: key stages 1 and 2 National curriculum in England*. Retrieved from: https://www.gov.uk/government/publications/national-curriculum-in-england-music-programmes-of-study [21.01.2019].

Department for Education (DfE) (2013c). *Music programmes of study: key stage 3 National curriculum in England*. Retrieved from: https://www.gov.uk/government/publications/national-curriculum-in-england-music-programmes-of-study [21.01.2019].

Department for Education (DfE)/Qualifications and Curriculum Authority (QCA) (1999). *Music: National Curriculum for England*. London: Her Majesty's Stationery Office (HMSO).

Department for Education and Skills (DfES) (2002). *Key Stage 3 National Strategy: Designing the Key Stage 3 Curriculum* (DfES 0003/2002). Annesley: DfES.

Department of Education and Science (DES) (1991). *National Curriculum: Music Working Group Interim Report*. London: DES.

Edexcel (2000). *Specification: Edexcel GCSE in Music (1426).* Mansfield: Edexcel.

Edmodo (2019) *The World's Largest Learning Community*. Retrieved from: https://www.edmodo.com/ [27.01.2019].

Fautley, M. (2010). *Assessment in Music Education*. Oxford: Oxford University Press.

Fautley, M. (2017). The Location of Creativity in Curriculum and Assessment in the Lower Secondary School in England. In R. Girdzijauskiene & M. Stakelum (Eds.), *European Perspective on Music Education: Creativity and Innovation* (pp. 71-83). Innsbruck: Helbling.

Fautley, M. & Colwell, R. (2018). Assessment in the Secondary School Classroom, in G.E. McPherson & G.E. Welch (Eds.), *Music Learning and Teaching in Infancy, Childhood, and Adolescence: An Oxford Handbook of Music Education, (vol. 2*) (pp. 257-276). New York / Oxford: Oxford University Press.

Fautley, M. & Daubney, A. (2015). *The National Curriculum for Music: A Framework for Curriculum, Pedagogy and Assessment in Key Stage 3 Music.* Retrieved from: http://www.ism.org/images/files/ISM_A_Framework_for_Curriculum,_Pedagogy_and_Assessment_KS3_Music_WEB.pdf [21.01.2019].

Fautley, M. & Savage, J. (2008). *Assessment of Composing at Key Stages 3 and 4 in English Secondary Schools: Research Report for the Associated Board of the Royal Schools of Music*. Retrieved from: http://www.jsavage.org.uk/jsorg/wp-content/uploads/2011/03/Assessment-of-Composing-at-Key-Stages-3-and-4-in-English-secondary-Schools-FINAL.pdf [26.01.2019].

Fautley, M. & Savage J. (2011). Assessment of composing in the lower secondary school in the English National Curriculum, *British Journal of Music Education*, *28*(1), 51-67.

Finney, J. (2011). *Music Education in England, 1950-2010: The child centred progressive tradition.* Surrey: Ashgate Publication Limited.

Finney, J. (2016). The place of music in the secondary school: ideology – history – justification. In C. Cooke, K. Evans, C. Philpott & G. Spruce (Eds.), *Learning to Teach Music in the Secondary School, (vol.2)* (pp. 4-16). New York/Abingdon: Routledge.

Firefly (undated). *The easiest to use school learning tool*. Retrieved from: https://fireflylearning.com/ [27.01.2019].

Gall, M. (2013). Trainee teachers' perceptions: Factors that constrain the use of music technology in teaching placements, *Journal of Music, Technology & Education, 6*(1), 5-27.

Gall, M. (2015). *Minutes of Postgraduate Certificate in Education Music School Mentor Meeting*. Bristol: University of Bristol.

Gall, M. & Breeze, N. (2005). Music Composition Lessons: the multimodal affordances of technology. *Educational Review*, *57*(4), 415-433.

Gall, M., & Breeze, N. (2007). The sub-culture of music and ICT in the classroom, *Technology, Pedagogy and Education*, *16*(1), 41–56.

Garvis, S. (2013). Beginning Generalist Teacher Self-efficacy for Music Compared with Maths and English, *British Journal of Music Education, 30*(1), 85-101.

Harrison, C. & Finney, J. (Eds.) (2010). *Finding The Student Voice*. Matlock, Derbyshire: National Association of Music Educators (NAfME).

Henley, D. / Department for Digital, Culture, Media and Sport (DCMS) (2011). *Music education in England*. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175432/DFE-00011-2011.pdf [26 .01.2019].

Kempa, R. & L'Odiaga, J. (1984). Criterion-referenced interpretation of examination grades, *Educational Research*, *26*(1), 56-64.

Kordes, U., Kafol, B. S. & Brunauer, A. H. (2014). A Model of Formative Assessment in Music Education, *Athens Journal of Education, 1*(4), 295-307.

Kwami, R.M. (2001). Music education in a new millennium, in A. Loveless and V. Ellis (Eds.), *ICT, Pedagogy and Curriulum: Subject to Change* (pp. 216-228). London/Abingdon/New York: Routledge-Falmer.

Mansell, W. (2007). *Education by Numbers: The Tyranny of Testing*. London: Politico's.

McClean, I. (2016). *Assessment at KS3*. Lecture to Music Trainee Teachers at the University of Bristol, 22 April 2016.

McIntosh, J. (2015). *Final Report of the Commission on Assessment without Levels*. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf [21.02.2019].

Mills, J. (2005). *Music in the school*. Oxford: Oxford University Press.

Mills, J. & Murray, A. (2000). Music technology inspected: good teaching in Key Stage 3, *British Journal of Music Education*, *17*(2), 129-156.

National Association for Head Teachers (NAHT) (2014). *Report on the NAHT Commission for Assessment*. Retrieved from: https://www.stem.org.uk/system/files/elibrary-resources/2016/04/Assessment%20commission%20report%20document%20%282%29.pdf [21.01.2019].

Norris, R. (2018). *Creativity can be taught to anyone. So why are we leaving it to private schools? The Guardian*, 17 January 2018. Retrieved from: https://www.theguardian.com/commentisfree/2018/jan/17/creativity-private-schools-uk-creative-industries-state [21.01.2019].

Odena, O. (Ed.) (2012). *Musical Creativity: Insights from Music Education Research*. Surrey: Ashgate Publishing Limited.

Office for Standards in Education (Ofsted) (2004). *Ofsted Subject Reports 2002/3 Music in secondary schools (HMI 1981)*. Manchester: Ofsted.

Office for Standards in Education (Ofsted) (2009). *Making more of music: an evaluation of music in schools 2005-08.* Retrieved from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/413347/Music_in_schools_wider_still__and_wider.pdf [21.01.2019].

Office for Standards in Education (Ofsted) (2012a). *Music in Schools: wider still, and wider – Quality and inequality in music education 2008-11*. Retrieved from: https://www.gov.uk/government/publications/music-in-schools [21.01.2019].

Office for Standards in Education (Ofsted) (2012b). *Making a mark: art, craft and design education*. Manchester: Her Majesty's Stationery Office (HMSO).

Paynter, J. (1992). *Sound and Structure*. Cambridge: Cambridge University Press.

Pearson (2019a). *Qualifications: Music – Edexcel A Levels.* Retrieved from: https://qualifications.pearson.com/en/qualifications/edexcel-a-levels.html [21.01.2019].

Pearson (2019b). *Qualifications: Music – EdExcel GCSE Music (2016).* Retrieved from: https://qualifications.pearson.com/en/qualifications/edexcel-gcses/music-2016.html [21.01.2019].

Pearson (2019c). *Qualifications: Music – Overview*. https://qualifications.pearson.com/en/subjects/music.html [21.01.2019].

Pratt, G. & Stephens, J. (Eds.) (1995) *Teaching Music in the National Curriculum*. Oxford: Heinemann.

Salaman, W. (2008). Reflections on progress in musical education. *British Journal of Music Education*, *25*(3), 237-243.

SMSC (undated). *What is SMSC?* Retrieved from: https://www.smscqualitymark.org.uk/what-is-smsc/ [26.01.2019].

Spruce, G. (1996). Assessment in the Arts: issues of objectivity, in G. Spruce (Ed.), *Teaching Music* (pp. 168-185) London: Routledge.

Spruce, G. (2016). An integrated approach to lesson planning, in C. Cooke, K. Evans, C. Philpott & G. Spruce (Eds.), *Learning to Teach Music in the Secondary School, (vol.2)* (pp. 80-93). New York/Abingdon: Routledge.

Swanwick, K. (1988) *Music, Mind and Education*. London: Routledge.

The Guardian (2018). *The Guardian view on educational priorities: arts not grammars*. 11 May 2018. Retrieved from: https://www.theguardian.com/commentisfree/2018/may/11/the-guardian-view-on-educational-priorities-arts-not-grammars [21.01.2019].

Thorpe, V. (2012). Assessment rocks? The assessment of group composing for qualification, *Music Education Research*, *14*(4), 417-429.

Wise, S., Greenwood, J. & David, N. (2011). Teachers' use of digital technology in secondary musiceducation: illustrations of changing classrooms, *British Journal of Music Education*, *28*(2), 117-134.

Appendix 1: National Curriculum for Music

(First introduced September 2013)(DfE, 2013b & 2013c)

Purpose of study

Music is a universal language that embodies one of the highest forms of creativity. A high-quality music education should engage and inspire pupils to develop a love of music and their talent as musicians, and so increase their self-confidence, creativity and sense of achievement. As pupils progress, they should develop a critical engagement with music, allowing them to compose, and to listen with discrimination to the best in the musical canon.

Aims

The national curriculum for music aims to ensure that all pupils:

- perform, listen to, review and evaluate music across a range of historical periods, genres, styles and traditions, including the works of the great composers and musicians
- learn to sing and to use their voices, to create and compose music on their own and with others, have the opportunity to learn a musical instrument, use technology appropriately and have the opportunity to progress to the next level of musical excellence
- understand and explore how music is created, produced and communicated, including through the inter-related dimensions: pitch, duration, dynamics, tempo, timbre, texture, structure and appropriate musical notations.

Attainment targets

By the end of each key stage, pupils are expected to know, apply and understand the matters, skills and processes specified in the relevant programme of study.

Key stage 1

Pupils should:

- use their voices expressively and creatively by singing songs and speaking chants and rhymes
- play tuned and untuned instruments musically
- listen with concentration and understanding to a range of high-quality live and recorded music
- experiment with, create, select and combine sounds using the inter-related dimensions of music.

Key Stage 2

Pupils should be taught to sing and play musically with increasing confidence and control.

They should develop an understanding of musical composition, organising and manipulating ideas within musical structures and reproducing sounds from aural memory.

Pupils should be taught to:

- play and perform in solo and ensemble contexts, using their voices and playing musical instruments with increasing accuracy, fluency, control and expression
- improvise and compose music for a range of purposes using the interrelated dimensions of music
- listen with attention to detail and recall sounds with increasing aural memory
- use and understand staff and other musical notations
- appreciate and understand a wide range of high-quality live and recorded music drawn from different traditions and from great composers and musicians
- develop an understanding of the history of music.

Key Stage 3

Pupils should build on their previous knowledge and skills through performing, composing and listening.

They should develop their vocal and/or instrumental fluency, accuracy and expressiveness, and understand musical structures, styles, genres and traditions, identifying the expressive use of musical dimensions.

They should listen with increasing discrimination and awareness to inform their practice as musicians.

They should use technologies appropriately and appreciate and understand a wide range of musical contexts and styles.

Pupils should be taught to:

- play and perform confidently in a range of solo and ensemble contexts using their voice, playing instruments musically, fluently and with accuracy and expression
- improvise and compose; and extend and develop musical ideas by drawing on a range of musical structures, styles, genres and traditions
- use staff and other relevant notations appropriately and accurately in a range of musical styles, genres and traditions
- identify and use the interrelated dimensions of music expressively and with increasing sophistication, including use of tonalities, different types of scales and other musical devices
- listen with increasing discrimination to a wide range of music from great composers and musicians
- develop a deepening understanding of the music that they perform and to which they listen, and its history.

Appendix 2: Attainment Targets for Music Prior to 2014 (Taken from DfE/QCA, 1999, pp.36-38)

Assessing attainment at the end of a key stage

In deciding on a pupil's level of attainment at the end of a key stage, teachers should judge which description best fits the pupil's performance. When doing so, each description should be considered alongside descriptions for adjacent levels.

Range of levels within which the great majority of pupils are expected to work		Expected attainment for the majority of pupils at the end of the key stage	
Key Stage 1	1-3	Age 7	2
Key Stage 2	2-5	Age 11	4
Key Stage 3	3-7	Age 14	5/6

Level 1

Pupils recognize and explore how sounds can be made and changed. They use their voices in different ways such as speaking, singing and chanting, and perform with awareness of others. They repeat short rhythmic and melodic patterns and create and choose sounds in response to given starting points. They respond to different moods in music and recognize well-defined changes in sounds, identify simple repeated patterns and take account of musical instructions.

Level 2

Pupils recognize and explore how sounds can be organized. They sing with a sense of the shape of the melody, and perform simple patterns and accompaniments keeping to a steady pulse. They choose carefully and order sounds within simple structures such as beginning, middle, end, and in response to

given starting points. They represent sounds with symbols and recognize how the musical elements can be used to create different moods and effects. They improve their own work.

Level 3

Pupils recognize and explore the ways sounds can be combined and used expressively. They sing in tune with expression and perform rhythmically simple parts that use a limited range of notes. They improvise repeated patterns and combine several layers of sound with awareness of the combined effect. They recognize how the different musical elements are combined and used expressively and make improvements to their own work, commenting on the intended effect.

Level 4

Pupils identify and explore the relationship between sounds and how music reflects different intentions. While performing by ear and from simple notations they maintain their own part with awareness of how the different parts fit together and the need to achieve an overall effect. They improvise melodic and rhythmic phrases as part of a group performance and compose by developing ideas within musical structures. They describe, compare and evaluate different kinds of music using an appropriate musical vocabulary. They suggest improvements to their own and others' work, commenting on how intentions have been achieved.

Level 5

Pupils identify and explore musical devices and how music reflects time and place. They perform significant parts from memory and from notations with awareness of their own contribution such as:

- leading others, taking a solo part and/or providing rhythmic support.

- They improvise melodic and rhythmic material within given structures, use a variety of notations and compose music for different occasions using appropriate musical devices such as melody, rhythms, chords and structures.
- They analyze and compare musical features.
- They evaluate how venue, occasion and purpose affects the way music is created, performed and heard. They refine and improve their work.

Level 6

Pupils identify and explore the different processes and contexts of selected musical genres and styles. They select and make expressive use of tempo, dynamics, phrasing and timbre. They make subtle adjustments to fit their own part within a group performance. They improvise and compose in different genres and styles, using harmonic and non-harmonic devices where relevant, sustaining and developing musical ideas and achieving different intended effects. They use relevant notations to plan, revise and refine material. They analyze, compare and evaluate how music reflects the contexts in which it is created, performed and heard. They make improvements to their own and others' work in the light of the chosen style.

Level 7

Pupils discriminate and explore musical conventions in, and influences on, selected genres, styles and traditions. They perform in different styles, making significant contributions to the ensemble and using relevant notations. They create coherent compositions drawing on internalised sounds and adapt, improvise, develop, extend and discard musical ideas within given and chosen musical structures, genres, styles and traditions. They evaluate, and make critical judgements about, the use of musical conventions and other

characteristics and how different contexts are reflected in their own and others' work.

Level 8

Pupils discriminate and exploit the characteristics and expressive potential of selected musical resources, genres, styles and traditions. They perform, improvise and compose extended compositions with a sense of direction and shape, both within melodic and rhythmic phrases and overall form. They explore different styles, genres and traditions, working by ear and by making accurate use of appropriate notations and both following and challenging conventions. They discriminate between musical styles, genres and traditions, commenting on the relationship between the music and its cultural context, making and justifying their own judgements.

Exceptional performance

Pupils discriminate and develop different interpretations. They express their own ideas and feelings in a developing personal style exploiting instrumental and/or vocal possibilities. They give convincing performances and demonstrate empathy with other performers. They produce compositions that demonstrate a coherent development of musical ideas, consistency of style and a degree of individuality. They discriminate and comment on how and why changes occur within selected traditions including the particular contribution of significant performers and composers.

Appendix 3

Class Formative Assessment Record Date

Name	Instrumental/ vocal experience/skills or special need / disability	Comments

Name	Instrument / SEN	Effort				Comments

Appendix

Formative Assessment Record — **Topic – Structure and Form (Tech)**

Name	Instrument / SEN	LAP/MAP /HAP	Ipad	Effort (VG = fully good, 1 = good, 2 = working)						Comments (? = task or Garageband Cloth, NU = Not understanding, D = distracted, M = melody)
				16/2	23/2	2/3	9/3	16/3	23/3	
[redacted]	K (ASD)	M		/	/	-	/	/	/	Slow to settle, not understanding – input skill. Tech issues, restarted. Began T1 (1/2)
[redacted]		H		/	/	-	/	VG	/	Struggle T1 (Student B not working well with him) finished T1+2 (working alone). Devel. A1 (Instr., loops)
[redacted]		-		/	/	-	/	/	/	Working on T1, finished T1+ almost 2. Devel. A1 (Instr., loops)
[redacted]	Guitar	H		-	/	-	/	/	/	Tech issues, finished T1+ almost 2. Devel. A1 (loops)
[redacted]	Drums	-		/	/	-	/	/	/	NU, just input skill. Tech issues, restarted on T1. T3
[redacted]		M	Y	-	-	-	/	/	/	Working with Student X. finished T1+ almost 2. Devel. A1 (loops)
[redacted]		H		/	/	-	/	VG	/	NU, just input Skill, finished T1. Devel. A1 (Instr., loops)
[redacted]	Piano, Drums	H		/	/	-	/	VG	/	Almost finished T1, finished T1,2+3. Devel. A1 (Instr., loops - appropriate
[redacted]	Singing	H		/	/	-	VG	/	VG	Almost finished T1, finished T1,2+3. Devel.A1 (Instr., loops – appropriate, helping others
[redacted]	Piano	H		/	/	-	/	/	/	¾ through T1. Tech issues, restarted. Devel. A1 (Instr., loops)
[redacted]		H		/	/	-	/	/	/	Good starter answers, T1, finished T1,2+3. Devel. A1 (Instr., loops) – appropriate
[redacted]		M		/	/	-	/	/	/	Tech problems, began T1. Finished T1+ almost 2. T3
[redacted]	Guitar	M	Y	/	/	-	/	/	/	Tech problems, began T1. finished T1+ almost 2. T3

Contributors

Dr. Andreas Breiter was appointed professor for Applied Computer Sciences in 2004 at Bremen University with an emphasis on the management of information and knowledge in education. He studied computer sciences, soiology and law at Goethe-University in Frankfurt/Main and received PhD in Computer Sciences in the year 2000 in Bremen. Andreas Breiter was Vice President for Research of Bremen University from 2017 until 2020 and is currently Acting Director of the Institute for Information Management Bremen (ifib).

Julia Finken holds a degree in informatics and is a member of the team of the Institute for Information Management Bremen (ifib). She was coordinator for the [PosyMus] Project.

Dr. Marina Gall is Senior Lecturer in Education for music at the University of Bristol. Prior to that she worked in Inner London Schools and in North Somerset. She has acted in various position for the European Association for Music in Schools (EAS). Among other duties she was PI in a project on school music and ICT in Europe.

Dr. Johannes Hasselhorn is currently researcher at the Faculty of Rehabilitation Sciences at the Technical University Dortmund. He received his Doctoral Degree in 2014 from Würzburg University of Music with a study on modelling practical competences in music. His emphasis is on quantitative methods in music education research whee he was involved in several projects ath the Universities in Dortmund, Bielefeld and Lübeck.

Dr. Jens Knigge studied music education and acred music. He earne a Doctoral Degree from Bremen University in 2010. He taught in music education programs in Stuttgart, Erfurt and Lübeck and is currently Full Professor at Nord University in Norway. Jens Knigge is an accomplished scholar applying mostly quantitative methods.

Dr. Andreas Lehmann-Wermser majored in music education, German litrature and educational science before starting a teaching career of two decades. After receiving a PhD he was appointed assistant professor at Bremen University. In 2015 he came Director of the Institute for Music Education research at the Hannover University for Music Drama and Media. He is Principal Investigator in several third party funded research projects on learning processes and social justice.

Benjamin Weyel studied music and media in Marburg and Paderbornand received his MA in 2011. He has worked in the field of Popular Music and Media in several institutions. Since 2015 he is assistant researcher in the Institute for Music Education Research in Hannover focusing digital technology and learning in music.

hmtmh Hochschule für Musik Theater und Medien Hannover [Institut für musikpädagogische Forschung]

[PosyMus] Potenziale von Feedbacksystemen im Musikunterricht

ifib Institut für Informationsmanagement Bremen GmbH

Symposium on e-assessment and electronic feedback in music lesson.

Friday, 15th September, 2017 / 9am – 4pm / Hanover University of Music, Drama & Media, Seelhorststraße 3, 30175 Hannover

Schedule

09.00 – 09.30	**Welcoming**
09.30 – 09.45	**Short presentation of „PosyMus" research project**
09.45 – 10.30	**On the value of Data-based Decision Making in education** Prof. Dr. Adrie J. Visscher, University of Twente
10.30 – 11.15	**PosyMus – the technical perspective** Institute for Information Management Bremen GmbH (ifib)
11.15 – 11.30	*Coffee break*
11.30 – 12.15	**Technology Based Competency Assessment in Music Education: the KOPRA-M and KoMus Tests** Prof. Dr. Jens Knigge, University of Music Lübeck Dr. Johannes Hasselhorn, TU Dortmund University
12.15 – 13.15	*Lunch break*
13.15 – 14.00	**Assessment in School Music: Problems and Possibilities** Prof. Dr. Marina Gall, University of Bristol
14.00 – 14.15	*Coffee break*
14.15 – 15.00	**PosyMus – the music educational perspective** Hanover University of Music, Drama & Media
15.00 – 16.00	**Discussion**

gefördert vom Bundesministerium für Bildung und Forschung

Appendix: Schedule of the Symposium

Digital media are of ever increasing importance in education. Music seems aspecially apt because they are essential in production and reception of music anyway. However, there is only limited research on their use in the classroom. The German Federal Ministry of Research and Education funded the three-year research program *PosyMus* (i.e. (potentials of systematic feedback in music classes“). It explored the field in a cooperation of information scientists and music educators. This volume documents the final international symposium that concluded the project.

Digitale Medien spielen in der Bildung eine immer größere Rolle. Musik bietet sich als Fach dafür besonders an, weil in Produktion und Rezeption diese Medien eine große Rolle spielen. Deren Einsatz ist freilich nur unzureichend erforscht. Das vom BMBF geförderte Forschungsprojekt *PosyMus* hat die Potenziale systematischen Feedbacks in Musik in der Zusammenarbeit von Musikdidaktikern und Informatikerinnen erkundet. Dieser Band dokumentiert das abschließende internationale Symposium.

ISSN 1430-8088
ISBN 978-3-931852-51-1
9 783931 852511
(D)29,80 €

hmtmh
Hochschule für Musik
Theater und Medien Hannover
Institut für musikpädagogische Forschung