Music Education: Stimulating Musical Creativity in Primary School Children through Non-Verbal Autonomy Support

R. D. Visser

Master Thesis – Talent Development & Creativity

[*S3212173*]
[*January*] [*2022*]
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
H.W. Steenbeek/L.H. Hendriks

Abstract

The present study aims to provide insight on the stimulation of musical creativity in primary education through an intervention on non-verbal autonomy support in primary school teachers and the subsequent musical creativity, characterized through convergent thinking and acting, of the children. Based on the literature that autonomy support stimulates creativity, and the idea that non-verbal autonomy support, such as gesturing, has additional value on top of verbal autonomy support, it is hypothesized that non-verbal autonomy support by teachers could enhance creativity in primary school children. Monte Carlo analyses and moving maximums were used to analyze the increases of non-verbal autonomy support in teachers. The Monte Carlo analyses were also used to assess increases in convergent thinking and acting in the children. Finally, we expected to see increases in non-verbal autonomy support in teachers and in convergent thinking and acting in children, and that these increases are correlated. State space grids were used to study the latter. Results showed that the intervention led to meaningful improvement in non-verbal autonomy support in teachers. Contrary to the expectation, results showed no support for improvement in convergent thinking and acting in the children. When we looked at the interaction of the former, the teacher-class combinations showed mixed results, as there were signs of coherence, but not always in a positive direction for both variables. Therefore, this research shows that primary school teachers can be trained to use non-verbal autonomy support, but that it does not directly lead to increased convergent thinking and acting in children.

Keywords: music education, non-verbal autonomy support, musical creativity, primary school, teacher intervention

Introduction

Music has played a role long before we spoke of human civilization. The first evidence of human involvement in (playing) music is estimated to origin from as early as 43,000 BC (Montagu, 2017). This evidence, however, only indicated use of musical instruments. Musical expression like dancing or singing may have existed even longer ago. Musical expression by humans develops dynamically. Whereas a few centuries ago the leading instruments in Western Europe were violins, harpsichords, and flutes (depending on the region), played by the likes of Vivaldi, Mozart, Bach, and Handel (e.g., Harthan, 1943; Marshall, 2005), recent musical expression is performed more digitally through the emergence of, for example, synthpop, house, and electronic dance music (EDM). Some (subjectively proclaimed) pioneers in these fields are Kraftwerk, Gary Numan, Daft Punk, and Skrillex (Castillo, et al., 2014; ‘Evolutie van elektronische muziek - Deel 1/3’, 2020; ‘Evolutie van elektronische muziek - Deel 2/3’, 2020; ‘Evolutie van elektronische muziek - Deel 3/3’, 2020). Not only has music developed over the years, it has also gotten more present in our everyday lives. Whether one is doing groceries in the supermarket or awaiting a response of a support line on the phone, it is highly likely that music is involved. Moreover, music has gotten more accessible over the last few decades. Whereas in the past one had to physically purchase an album at a music store, now, with the emergence of streaming platforms like Spotify, Deezer, and Apple Music, music is more accessible than ever. As one can tell, music develops continuously in many facets at a macro level, and it goes hand in hand with continuous innovation. Innovation that might be facilitated by the purposeful intention of music performers to look for ways to communicate and express themselves through music (Juslin, 2003). Therefore, it is important that children get accustomed to (the creation of) music from an early age and learn the importance of music.

Music education and the role of creativity

Aróstegui (2016) says that the point of attention has been too much on the implementation of music in education and not so much on how it should be implemented. There has been too much of an emphasis on music education as a practice rather than as research, he states. This indicates a will to implement music education in the educational system, but there is not enough knowledge on how to do that (yet). Moreover, he concluded that there is a world-wide decline in music education. To counteract this problem, there have been established some initiatives already. In the Netherlands for example, *Muziek in de Klas* [Music in the class] is an initiative aimed at implementing more music lessons in primary education. But this again, is mainly focused on the need to implement music in education and its benefits, rather than on how it should be implemented.

Just as music development takes place in the broader sense within society, it also happens intrapersonal. A crucial factor in the dynamic process of music development within people, and therefore also in children, is creativity (Webster, 2002). The milestones and inventions in produced music would have never happened without the creativity of some explorative individuals. In contrast, music education focused on creativity is slow in its development, because music research has been lacking in their emphasis on creativity and creative stimulation (Webster, 2009). Creativity is often described as a matter of originality or novelty within an area in such a way that it is also appropriate for that specific context (e.g., De Dreu et al., 2008; Malinin, 2019). Webster (1990) describes two components of creativity in music: divergent and convergent thinking. He explains convergent thinking as a factual approach to interacting with music. That is, being able to use various kinds of musical components together to form a satisfying final product. Divergent thinking, he claims, is independent from this product approach. He describes it as a more explorative way of thinking about music, namely: exploring the possibilities offered by the internal musical data bank of the child. It is the complementation of both convergence and divergence that causes a creative creation to emerge (Yusof & Tan, 2018). Moreover, both convergent and divergent thinking and acting are agreed upon by researchers to be ordinary cognitive processes that contribute to creativity (Malinin, 2019).

Autonomy, autonomy support, and self-determination theory (SDT)

The question of Aróstegui (2016) on how music should be taught in education, is partly answered by some prior research. A recent review of Alves-Oliveira et al. (2021) on creativity tells us that the most used tool to get children acquainted with creativity is through verbal instructions on how to follow a program. However, apart from the intrapersonal development of music, an interpersonal element is also important. Kupers et al. (2018) state that creativity is embedded in the interaction between child, environment, and the child's social encounters, as part of a complex dynamic system. Subsequently, children in particular, are capable to intuitively take initiative in the creation of music (Kondo & Wiggins, 2018). Moreover, Kupers et al. (2018) state that the development of creativity is not static but dynamic. Van Vondel et al. (2017) and Turner and Christensen (2020), also emphasize the importance of teacher-student interaction. Moreover, Van Vondel et al. (2017) specifically studied the role of teacher instruction and the vital role it has in the learning dynamics of children. Thought-provoking, inquiry-based instructions, they say, lead to enhanced participation on a cognitive level by children. On top of that, Reeve and Cheon (2021) pose that autonomy supportive teaching can very well be introduced in the already existing teaching style used by a teacher. From the three psychological human needs (the need for autonomy, relatedness, and competence), the need for autonomy is crucial for stimulating intrinsic motivation, according to the self-determination theory (Deci & Ryan, 2000). Ryan and Deci (2000) pose that humans are proactive and engaged when the three needs of the self-determination theory are met. That is, when humans feel validated by others in their doing, get the freedom to autonomously explore, and get positive feedback on reasonably challenging activities, their intrinsic motivation to engage in that behavior increases. Supporting autonomy has benefits for multiple aspects, it leads, for example, to feelings of happiness and vitality when succeeding in a task while feeling autonomous (Ryan & Deci, 2001). Moreover, when in relation to others, autonomy support is positively related to variability in character traits like increased showing of openness, conscientiousness, agreeableness, and expressions of the self (La Guardia & Ryan, 2007).

Stefanou et al. (2004) pose autonomy support in the classroom to be a way of teaching, both personal and instructional, that enables the developing sense of autonomy in students. They made a distinction within autonomy support between organizational, procedural, and cognitive autonomy support. From these three, cognitive autonomy support shows the most promise in promoting deep-level thinking. Webster (1990) praises cognitive autonomy support for its promotion of deep-level thinking, but what he emphasizes as important elicitors for both divergent and convergent thinking (e.g., encouraging questions), corresponds very much with what Stefanou et al. (2004) describe as cognitive autonomy support as well. Concordantly, Webster (2002) poses that learning is more effective when the learner is involved actively rather than a passive receiver of information.

Learning in music education through non-verbal autonomy support

Music learning happens actively and socially (Kondo & Wiggins, 2018). Music education is particularly suited for this, as making music encompasses activity and social interaction by nature (Bishop, 2018). The latter corresponds with the enactive component of the 4E cognition perspective, while the relation to others is in line with how the 4E cognition perspective states that cognition is embedded in social interaction (Malinin, 2019). That is, the emergence of creativity extends beyond the individual through the creative potential that the body shapes in enactive interaction with both the environmental and social input. The enactive perspective also sees humans as autonomous systems that form their surroundings in a way that it provides meaning and value to that person. This is often described as creative experiences (Malinin, 2019). This is in line with how the complex dynamic systems approach states that learning emerges (e.g., Kupers et al., 2018; Van Geert, 2008). Moreover, many of the benefits of teacher autonomy support in a classroom described by Simon and Salanga (2021) align with components of the 4E cognition perspective, like improved inter- and intrapersonal factors, which may therefore, subsequently, support creativity. In the opinion of Webster (2002), school educators can play a pivotal role in the process of autonomy support in favor of creativity, as he sees education as a partnership between teacher and student in which the teacher has an architectural role. This is in line with the non-linear complex dynamic systems theory, which states that development takes place non-linearly through interaction of many interdependent factors (Van Geert, 2008), like the interaction between student and teacher (Steenbeek & van Geert, 2007). Within the development of a child, its surroundings, including the people it interacts with, are part of these factors. Therefore, the role of the teacher may very well be pivotal in the development of the child and being autonomy supportive is likely to have positive effects. Supporting autonomy in children can be done verbally but shows even more promise when accompanied by gestures (Novack & Goldin-Meadow, 2015). According to Simones (2017), gestures are also important in music education. Gestures by a teacher give direction to the learner, but because the gestures still need to be interpreted, they offer room for autonomous interpretation. Nonetheless, the scientific ground for the benefits of gesturing on (musical) creativity has not yet been established (Simones, 2017). Reeve and Cheon (2021) amplify the benefits of autonomy supportive teaching in supporting the intrinsic motivation of a child and its internalization of the material presented by the teacher. Correspondingly, this leads to the wide range of benefits priorly mentioned for both student, teacher, and the classroom climate, in educational settings.

Portraying dynamic development of convergent thinking and acting

The dynamic skill theory of Fischer and Bidell (2006) supports the further concretization of non-verbal autonomy support. They have delved more into how cognition and skill develop within activity. Their dynamic skill theory poses that cognitive development takes place in several tiers: Sensorimotor Actions, Representations and Abstractions. They subdivided these three tiers in a single, mapping and systems level. Expression of a developing skill gradually gets more complicated as a human progresses through these levels. Likewise, as convergent thinking and acting serve in the active creation of a musical product (Webster, 1990), this theory also serves the development of convergent creativity within music. When a systems level of a tier is reached, the door opens to the next tier. For example, when someone has reached the systems level of the actions tier, they now can take the step towards single representations. Noteworthy to mention is the fact that Fischer and Bidell (2006) pose development to take place dynamically. That is, through the developing process, it is possible to sometimes show a skill on an already accomplished level of a tier. Again, of importance is the role of others as one of the many living systems surrounding a developing child. The child might interact variably, depending on its surroundings. The expressed skill in front of its parents might be, for example, of a much higher level than in front of its classmates. Here again, the importance of supportive interactions is emphasized. Their dynamic skill theory encompasses a scale that makes it possible to visualize the dynamic interaction and, therefore, provides a way to study this on a micro timescale. Subsequently, different micro timescales can be compared to study differences on different time intervals.

To summarize: both convergent and divergent thinking and acting are complementary cognitive processes of creativity that can emerge through active and enactive interaction with the environment. These two processes, in their turn, are possibly stimulated through autonomy support, which can be done both verbally and non-verbally. On which teachers can be taught to have positive meaningful impact by scaffolding children to a higher level. The question arises if non-verbal autonomy support leads to more creative expression, as it shows promise for supporting autonomy (Novack & Goldin-Meadow, 2015) but has not been linked yet to musical creativity (Simones, 2017).

Present study and research questions

This research tries to contribute to answering the before mentioned question as part of the Curious Minds project on music education (Hendriks, Steenbeek & Bisschop Boele, 2018), which aims to use the natural curious characteristics of children in motivating their learning (Van Benthem et al., 2005). This is done by posing an intervention on teachers within music lessons on primary schools. Therefore, the question is: "What is the influence of a primary school teacher intervention (as part of the Curious Minds project) on teachers' non-verbal autonomy support and primary school students' convergent thinking and acting in musical lessons?" Based on the prior literature, we hypothesize that expressed non-verbal autonomy support by the teacher can be increased through coaching, that we will see an increase in convergent thinking and acting in primary school children in music lessons, and that both prior increases are correlated.

Method

Participants

The study was conducted in four primary schools in the north of the Netherlands. From these schools, six teachers (*N* = 6, 6 female) were video recorded during a maximum of 10 music lessons. 2 lessons before, 4 lessons during, and 2 lessons after the intervention. The teachers had an average age of 36 (SD = 4.0). This resulted in a pre-measurement (PRE), intervention (INTER), and post-measurement (POST) period. Post-measurement took place between 4 weeks and 5 months (due to the Covid-19 pandemic) after intervention. The classes could be any of the Dutch primary school middle classes (group 3 to 6), which children normally progress through from 4 to 12 years of age. The age range of participating children went from 6 to 10 years old. Class sizes were not smaller than 19 and not bigger than 25 children (M = 23, SD = 1.9).

Design

Recordings were made with two cameras. Parents of all children were approached to ask for consent for video recordings. Children whose parents did not comply were not shown on camera and therefore not included in the research. One camera was focusing on the teacher and the class as a whole, the other on students in the classroom whose parents gave active permission. There were ten lessons recorded from all but one teacher, of which the last recording could not take place due to the Covid-19 pandemic. From each recorded lesson, four fragments were chosen with a combined duration of 10 minutes: 2 minutes from the introduction; 6 minutes from the middle part in which teacher and children were actively working on musical tasks; and 2 minutes from the last part of the lesson where children often showed their final musical pieces. The video recordings were coded on non-verbal autonomy support (ASNV) in the teacher and convergent thinking and acting (CTA) in the children with mediacoder.gmw.rug.nl (Bos & Steenbeek, 2009).

Procedure

First, two lessons were recorded. After the second lesson, the teachers were introduced to pedagogical-didactic strategies involving an autonomy supportive teaching style and stimulation of musical creativity, and practice-based exercises. This was the start of the intervention in which all teachers individually were coached and given video feedback (for more information about the intervention, see Hendriks et al., 2018). For each of the following four lessons (lessons three to six), a video feedback session of an hour was provided afterwards. The intervention was performed by the main researcher and a teacher of the department of music education of the conservatoire. Lesson seven and eight were taken as post-measurement. The delay between lesson six and seven differed between the teachers. For three teachers the seventh lesson was recorded two to three weeks after the last intervention, and the eighth lesson after four to six weeks after the last intervention. For two teachers the post-measurement took place 5 months after the last intervention. This was because of the Covid-19 outbreak and the subsequent closing of the schools and limited allowed physical visits as precautions taken by the Dutch government. Not all teachers were recorded for the last lesson of the post-measurement, as for one teacher there is no recording of the tenth lesson.

During PRE, the music lessons were constructed in many ways. Teachers were told to teach “as usual”. This means that sometimes the lessons will be aimed at singing and the musical expressions of the children will be more vocally. Other lessons will be more rhythmic based or instrument based. During INTER, the teachers were asked to give lessons suited for musical creativity. Therefore, rhythmic based, imitation based (e.g., weather), or body percussion were more prevalent during these lessons. The content of POST lessons was not based on any specific instructions.

Variables

The coding of ASNV in the teachers and CTA of the children was according to a coding scheme (Appendix A and B), adapted from Van Vondel et al. (2017). First, a video was watched and occurrences of expression (ASNV or CTA) were marked. Then, the video was watched again and every mark was coded on a specific level of ASNV or CTA. Different expressions of ASNV by the teachers were ranked on a 1-to-8 ordinal scale, which measured the extend of ASNV, in which 1 was a stop-sign and 8 was encouragement (Appendix A). Coding of CTA had a 1-to-9 scale with a build-up similar to the dynamic skill theory of Fischer and Bidell (2009), which has been used in previous research (e.g., Van Vondel et al., 2017) and was now adapted for use in primary music education. This means that the nine different codes were subdivided in three tiers: actions, representations, and abstractions. Each of these three tiers has the same three sub-levels: single, mapping and systems. For a child to reach the single level of one of the tiers, it usually first must have achieved the systems level of the prior tier. This means that a child first surpasses, for example, the systems representations level before reaching the single abstractions level (Appendix B). The different codes are further explained in more detail in Appendix B. For the reliability of both the coding systems, a Cohen's kappa (Viera & Garrett, 2005) has been calculated over 15% of the data for the interrater reliability between the main researcher and the master student(s). This resulted in k = .78 for ASNV and k = .69 for CTA.

Analysis and statistical tools and figures

Development of ASNV through video feedback (RQ1)

The first research question, whether teachers improved in ASNV during and after the intervention, was answered in three ways. (1) First, we compare the frequencies of each code in each research phase. This gives us a first impression on the overall performances on ASNV during the research. (2) Then, the average ASNV score for INTER and POST, in comparison with that of PRE, were subjected to a Monte Carlo (MC) analysis (Todman & Dugard, 2001). This way, despite the relatively small *N* for the teachers, we can estimate the probability of the presence of ASNV in the future, thanks to 1000 shuffled permutations based on the raw data, also aiding in finding the probability of the mean differences in the results. (3) To give a perception of the fluctuations and the height of ASNV that teachers expressed, we created a moving maximum graph (MM) of the development of a teacher over the time frame of 8 lessons. This way, an image of ASNV in all three research phases could be made, visualizing the progression that these teachers possibly made. These three analyses aid us in making conclusions on group level. The window of a time frame in the MM was 60 seconds for the whole length of eight times the 600 seconds per lesson. The last 60 seconds are not included, as the limit of the dataset causes the window to get smaller near the end, possibly leading to disturbed interpretations. In order to plot the MM, we adjusted every code in the coding scheme bigger than 5, one upwards. All the zeroes (i.e., moments without ASNV expression) are replaced with fives, therefore creating a neutral level with low and medium levels (1-4) on one side and medium-high and high levels (6-9) on the other. All values above five are seen as autonomy supporting non-verbal acts by the teacher. We also counted the peaks, which we define as reaching 8 or 9 (high ASNV) in the graph (code 7 and 8 in the coding scheme). The peak ends when it gets below 8 again. This resulted in a visualization of the maximum ASNV reached by the teachers. MM of teachers 7, 9, and 10 are presented as representative cases of different developmental trajectories. Therefore, together, they provide a broad image of possible developments. The MM of the other teachers can be found in Appendix C.

To draw a final conclusion on the training effect of individual teachers in ASNV, we label higher levels of ASNV as 'meaningful' when the difference is significant ($p < .05$) and Cohen's d_s is at least medium (> 0.5; Lakens, 2013), 'slightly meaningful' when the *p* value is between 0.05 and 0.01 with Cohen's d_s between 0.2 and 0.5, and 'not meaningful' when Cohen's d_s is below 0.2, regardless of the *p* value. On a group level, we then conclude whether we accept the hypothesis (at least three meaningful increases and one slightly meaningful increase in teacher ASNV), partly accept the hypothesis (at least two meaningful increases *or* more than three slightly meaningful increases), or whether we reject the hypothesis (fewer than two meaningful increases in teacher ASNV).

Development of CTA (RQ2)

In similar fashion as with ASNV, (1) frequencies of each code over all research phases gives us a first impression of the development of CTA for all classes. (2) For visualization of the trajectory of all classes the average scores per lesson are shown in a graph. To find out whether the progression in CTA during the experiment was significant, we administered the same procedure as with ASNV in teachers. (3) That is, we performed a MC analysis for the mean differences between PRE and INTER, and PRE and POST for all classes. All these analyses aid us in making conclusions on group level.

Just like with ASNV, we label higher levels of CTA as ‘meaningful’ when the difference is significant ($p < .05$) and Cohen’s d_s is at least medium (> 0.5), ‘slightly meaningful’ when the *p* value is between 0.05 and 0.01 with Cohen’s d_s between 0.2 and 0.5, and ‘not meaningful’ when Cohen’s d_s is below 0.2, regardless of the *p* value. On a group level, we again conclude whether we accept the hypothesis (at least three meaningful increases and one slightly meaningful increase in CTA), partly accept the hypothesis (at least two meaningful increases *or* more than three slightly meaningful increases), or whether we reject the hypothesis (fewer than two meaningful increases in class CTA).

Relationship between ASNV and CTA (RQ3)

For a first indication of the coherence of ASNV and CTA, a Spearman’s rho correlation is calculated in Microsoft Excel for each class individually. Meaning, both ASNV and CTA averages were ranked, and the correlation was calculated subsequently. Spearman’s rho is chosen because the data is non-parametric, non-linear, and measured on an ordinal scale (Puth, Neuhäuser, & Ruxton, 2015).

Secondly, a State Space Grid (SSG; e.g., Turner & Christensen, 2020) is formed for the time series of ASNV and CTA, to be able to make a statement on their interdependent coherence (i.e., does class CTA follow teacher ASNV?). ASNV was depicted on the x-axis and CTA on the y-axis. We wanted to analyze the mean number of events for combinations of high ANSV and CTA, and for combinations of low ASNV and CTA. Hence, we divided the grid in four regions. Region 1 (R1) would be in the upper right corner of the grid (ASNV 5 to 8, CTA 6 to 9). The line was drawn at 5 for ASNV, because this is the first code in which the teacher acts to stimulate autonomy, and at 6 for CTA, as this is the first level at which the musical play of children sounds as a musical whole and is played convincing. Region 3 (R3) is the down left corner of the grid (ASNV 1 to 4, CTA 1 to 5). For INTER and POST, compared to PRE, we expected to see an increase of events in R1, and a decrease of events in R3. Significance was again calculated with the MC analysis. To compute effect size, we used Cohen's d_s which accounts for differences in group sizes (Lakens, 2013). Furthermore, we used the SSGs to qualitatively analyze the trajectories of teachers. We chose to show one prime example of what we hoped to see (teacher-class combination 10), and one that contrasts with our hypothesis (combination 8). The remaining SSGs can be found in Appendix D.

A positive development in the relation of ASNV and CTA for a teacher-class combination is called 'meaningful' when the correlation between them is at least moderate positive (strong: $\rho > .8$, moderate: $.8 > \rho > .4$; 'Numeracy, Maths and Statistics - Academic Skills Kit', n.d.) *and* the mean number of events increased for R1 and decreased for R3 (for at least one of INTER or POST). When the correlation is positive ($\rho > 0$) *and* an increase in events for R1 *or* a decrease for R3 is found (for at least one of INTER or POST), then we label it as 'slightly meaningful'. When only one of the three (Spearman's rho, increase in R1, decrease in R3) *or* when neither of the three is present, the coherence (if any) is 'not meaningful'.

Results

Development of ASNV through video feedback (RQ1)

Distribution (analysis 1, frequencies)

Over the three research phases, pre-measure, intervention, and post-measure, the distribution of the event-based codes for ASNV are shown in Table 1. This shows the percentage that each ASNV code was present during that phase of the research. This gives us an indication of overall performances of the teachers during the research phases. Remarkably, while only two teachers had a score of 8 on ASNV in the pre-measurement period, all teachers reached this height in the intervention period and 5 teachers still had scores of 8 in the post-measurement period. Code 4 is the most frequent throughout the whole research. Code 8 appeared the least during PRE, but during INTER and POST code 1 was the least frequent code, giving an indication of a shift in ASNV.

Table 1

Count of Codes for ASNV

Code	Pre-Measurement	Intervention	Post-Measurement
1	8 (1.02%)	6 (0.42%)	3 (0.46%)
2	145 (18.47%)	113 (7.86%)	56 (8.56%)
3	38 (4.84%)	140 (9.74%)	63 (9.63%)
4	304 (38.74%)	567 (39.43%)	225 (34.40%)
5	39 (4.97%)	88 (6.12%)	35 (5.35%)
6	175 (9.55%)	393 (27.33%)	171 (26.15%)
7	39 (4.97%)	69 (4.80%)	51 (7.80%)
8	4 (0.51%)	24 (1.67%)	8 (1.22%)
“O”	33 (4.20%)	38 (2.64%)	42 (6.42%)
Total	785 (100%)	1438 (100%)	654 (100%)

Individual teachers (analysis 2, MC analyses)

As can be seen in Table 2, all but teacher 7 showed significant increase in ASNV on INTER, compared to PRE. For POST, the MC analysis showed that ASNV scores were (still) significantly larger than during PRE, for all but teacher 11. For teachers 7, 8, and 10, the scores during POST were even higher than during INTER. When we look at the meaningfulness, five teachers showed significant increase in ASNV for INTER of which three were meaningful (Teacher 8, 9, and 10) and two slightly meaningful (Teacher 11 and 12), according to Cohen's d_s. For POST, also five teachers showed significant increase, this time two were meaningful (Teacher 8 and 10) and three were slightly meaningful (Teacher 7, 9, and 12). Noticeably, both teachers 8 and 10 had meaningful increases for INTER and POST, according to effect sizes.

Table 2

ASNV Scores per Teacher

	Pre-Measurement		Intervention				Post-Measurement			
Teacher	M	SD	M	SD	*p*	Cohen's d_s	M	SD	*p*	Cohen's d_s
7	4.2	1.5	4.1	1.6	.982	-0.10	4.7*	1.8	<.001	0.28
8	3.1	1.7	4.2**	1.7	<.001	0.66	4.9**	1.5	<.001	1.08
9	4.1	1.7	4.9**	1.6	.001	0.50	4.5*	1.5	<.001	0.22
10	3.5	1.5	4.4**	1.6	<.001	0.56	4.8**	1.5	<.001	0.79
11	4.0	1.7	4.7*	1.7	<.001	0.38	4.0	1.5	.571	-0.01
12	3.9	1.8	4.7*	1.6	<.001	0.46	4.7*	1.8	<.001	0.43

Note. *p*-values and Cohen's d_s's are measurement of comparison with the pre-measurement phase.

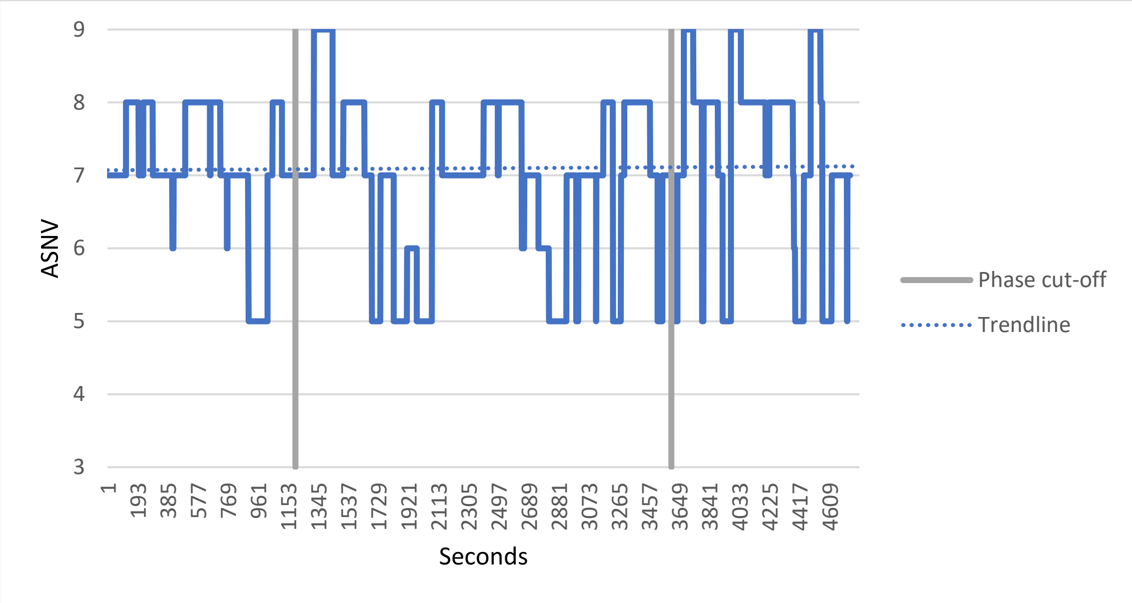
* Slightly meaningful, ** Meaningful.

Individual teachers (analysis 3, moving maximum)

Teacher 7. The MM of the development of this teacher is shown in Figure 1. Remarkable for this teacher is that during PRE, she already scores consistently high, also indicated by the intercept of the trendline of $y = 7.07$. Moreover, as can be seen from Figure 1, this teacher only began to show the highest level of ASNV during INTER and even showed this more often in POST. However, POST had as many peaks as PRE (both 5) but INTER had more (7).

Figure 1

Moving Maximum Teacher 7

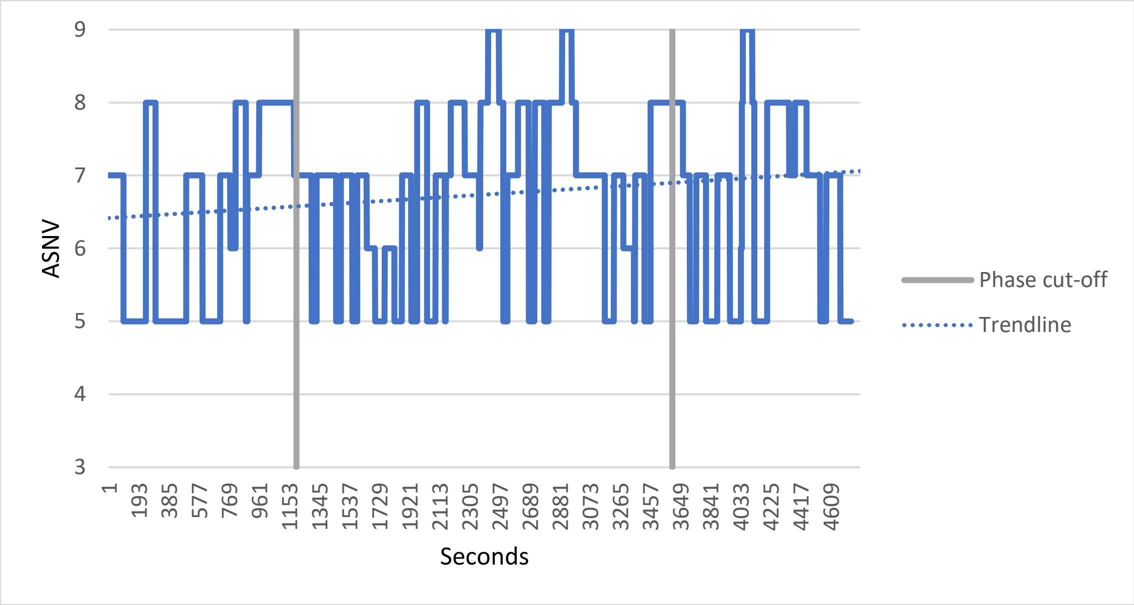


Note. This figure contains the moving maximum of teacher 7. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.00001x + 7.07$, with x in seconds.

Teacher 9. Figure 2 shows the moving maximum of teacher 9. With a lower intercept of $y = 6.42$ in PRE, this teacher had more room for improvement than teacher 7 (Figure 1). The slope was also a tenfold bigger than that of teacher 7. Therefore, the trendline is also steeper. There was also an increase in peaks for INTER (7) and POST (4) compared to PRE (3). Like teacher 7, this teacher only began to show the highest level of ASNV during and after INTER.

Figure 2

Moving Maximum Teacher 9

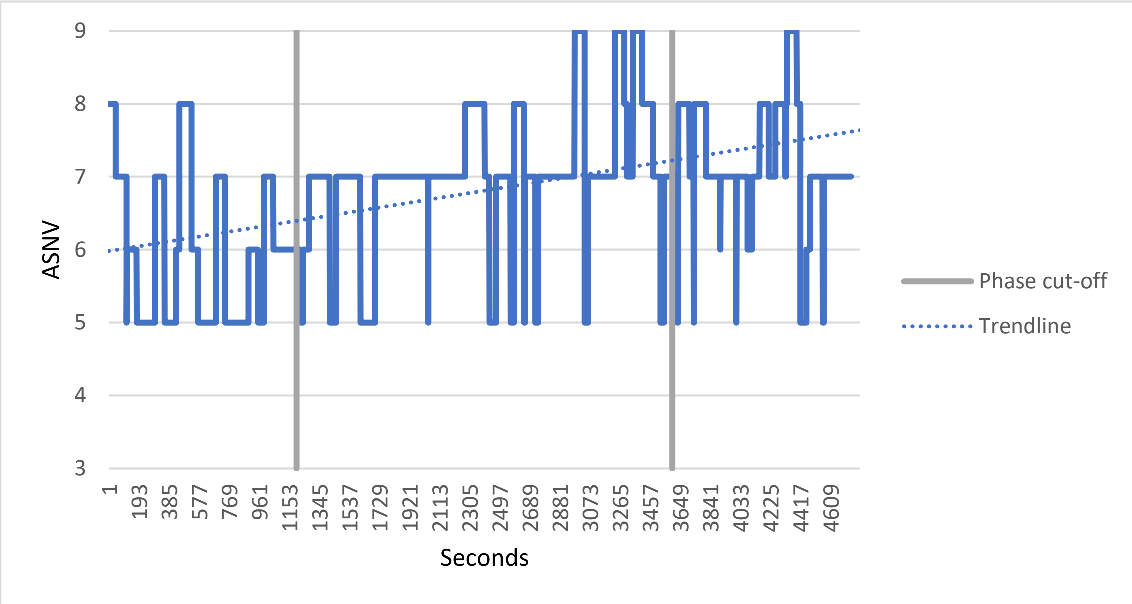


Note. This figure contains the moving maximum of teacher 9. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.0001x + 6.42$, with x in seconds.

Teacher 10. Figure 3 shows a pattern exemplary of what we hoped to see in this experiment. It clearly shows that teacher 10 sits increasingly less at, or below, the neutral 5 while time passes, while simultaneously spending more time at the higher ASNV levels. Moreover, an increase of peaks at the highest ASNV level is seen during INTER and POST (both 5 peaks) compared to PRE (2 peaks). The trendline of this teacher is also the steepest of all with a slope of 0.0003 units per second.

Figure 3

Moving Maximum Teacher 10



Note. This figure contains the moving maximum of teacher 10. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.0003x + 5.98$, with x in seconds.

Summarized results on ASNV

The shift in the most prevalent codes from PRE to INTER and POST, with more higher and fewer lower codes, is a first sign of support for the hypothesis. In addition, the significant higher averages in INTER and POST for five teachers (in both cases), compared to PRE, further supports the hypothesis. Overall, these former analyses combined with additional support from the MMs, provide strong support for the hypothesis that ASNV meaningfully increased during the study.

Development of CTA (RQ2)

Distribution (analysis 1, frequencies)

Over the span of the 12 lessons during the pre-measurement, 24 intervention lessons and 11 lessons of post-measurement, the distribution of the event-based scores of CTA is shown in Table 3. It shows each portion of the total distribution of that research phase per code and, therefore, gives us an indication of the development of CTA during the experiment. For example, we see an increase of code 6 for INTER and POST, which tells us that there was more coherent play in these research phases. Furthermore, we see an increase for code 2 and 3, and a decrease for code 7, 8, and 9 for INTER, but we see a decrease in code 2 and 3 and an increase in code 7, 8, and 9 for POST. Showing improvement for POST, but not for INTER. Code 5 was the most prevalent during all research phases.

Table 3

Descriptive Statistics for CTA

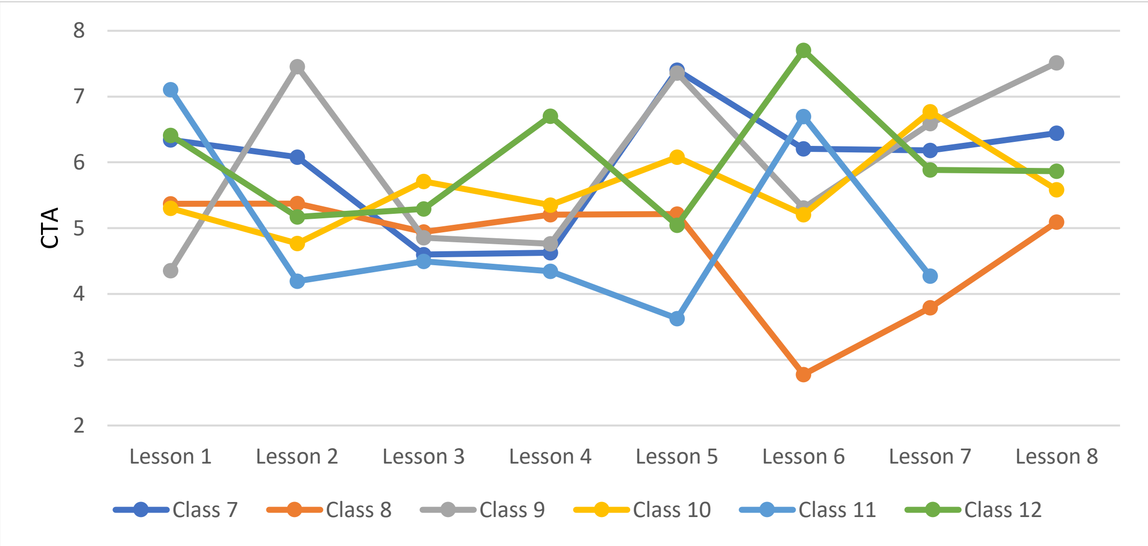
Code	Pre-Measurement	Intervention	Post-Measurement
1	0 (0%)	6 (1.08%)	0 (0%)
2	16 (6.75%)	50 (9.04%)	9 (3.40%)
3	36 (15.19%)	97 (17.54%)	18 (6.79%)
4	25 (10.55%)	99 (17.90%)	51 (19.25%)
5	69 (29.11%)	153 (27.67%)	73 (27.55%)
6	23 (9.70%)	67 (12.12%)	37 (13.96%)
7	11 (4.64%)	21 (3.80%)	26 (9.81%)
8	45 (18.99%)	43 (7.78%)	37 (13.96%)
9	12 (5.06%)	17 (3.07%)	14 (5.28%)
Total	237 (100%)	553 (100%)	265 (100%)

Individual classes (analysis 2, averages)

Figure 4 shows us the average score on CTA for each class per lesson. Here, except for the first two lessons of the intervention (lesson 3 and 4), it can be seen as a lot of variability. But when we look at class 10, for example, we see an upward trend. We also see that the class that started off the lowest, had the highest scores during POST. What also catches the eye is that all but class 8 showed relatively high scores during lesson 6. Class 8 had a very consistent pattern and then suddenly dropped at lesson 6.

Figure 4

Average CTA Score per Class



Individual classes (analysis 3, MC analyses)

Table 4 shows that only class 10 and 11 showed significant improvement in CTA during INTER compared to PRE, as analyzed through the MC analysis. Both increases were slightly meaningful, according to Cohen's d_s. The CTA score of class 10 remained bigger in POST compared to PRE. It even grew in comparison to INTER. Furthermore, besides class 10, class 9 had also a significantly higher CTA score during POST. Both increases were meaningful according to Cohen's d_s. All the other classes did not have meaningful increases in CTA. Moreover, just three classes improved upon PRE for INTER. This number was slightly bigger for POST, as four classes improved upon PRE in that period.

Table 4

CTA Scores per Class

	Pre-Measurement		Intervention				Post-Measurement			
Class	M	SD	M	SD	*p*	Cohen's d_s	M	SD	*p*	Cohen's d_s
7	6.2	1.9	5.8	1.9	1	-0.22	6.3	1.6	.159	0.05
8	5.4	1.8	5	1.2	1	-0.27	4.7	1	1	-0.51
9	6.2	1.9	5.8	1.5	.981	-0.23	7.2**	1.8	<.001	0.53
10	5.0	1.1	5.6*	1.6	<.001	0.39	6.2**	1.1	<.001	1.10
11	4.1	2.1	4.6*	1.7	<.001	0.23	4.3	1.1	.195	0.08
12	6.0	1.7	6.1	1.6	.283	0.04	5.9	2.1	.878	-0.08

Note. *p*-values and Cohen's d_s's are measurement of comparison with the pre-measurement phase.

* Slightly meaningful, ** meaningful.

Summarized results on CTA

The frequencies table (Table 3) gave some indication of growth in CTA, but only for the frequencies of codes during POST, with increases in the relative amount of code 6 and 7, and decreases of codes 2 and 3, compared to PRE. Based on these frequencies, strong support for the hypothesis is not yet provided. Figure 4 does not provide full support for the hypothesis either, as only some of the classes show a clear positive progression in average CTA scores. Moreover, the differences in average CTA scores per research phase, expressed in significance and meaningfulness, give support to the hypothesis only in the case of POST. Therefore, the hypothesis that CTA would increase during the study, is not supported.

Coherence between ASNV and CTA (RQ3)

General coherence (Spearman's correlations)

To get a first impression of the coherence of ASNV and CTA, correlations between the two variables for each teacher were calculated. Results are shown in Table 5. Two combinations have a moderate positive correlation (Combination 9), three have a weak positive correlation (.4 > ρ > 0; 'Numeracy, Maths and Statistics - Academic Skills Kit', n.d.) (Combination 10, 11, and 12), and one has a negative correlation (Combination 8). What catches the eye is that teacher 8 has a negative correlation, which would mean that her positive progression of ASNV is paired with a downslope of CTA in the musical expression of her class. Moreover, teacher 12 barely shows any correlation at all.

Table 5

Spearman Rho Correlation Coefficients of ASNV-CTA Combinations

Combination 7	Combination 8	Combination 9	Combination 10	Combination 11	Combination 12	Together
.45	-.81	.71	.31	.21	.01	.26

Note. Spearman's rho correlation for the ranked averages per lesson per teacher-class combination.

State Space Grids (mean number of events)

The mean number of events of R1 per lesson, for all teacher-class combinations, are shown in table 6 and the mean number of events of R3 are presented in table 7.

Table 6

Mean Number of Events in R1 of the SSGs for ASNV and CTA

	Pre-Measurement	Intervention		Post-Measurement	
	M	M	Cohen's d_s	M	Cohen's d_s
Combination 7	7.50	4	-1.52	12.50**	2.50
Combination 8	3	3.50	0.19	1.50	-1.29
Combination 9	4	4.75	0.19	5	0.36
Combination 10	2.50	4.50	0.92	5.50	3.31
Combination 11	6	2.50	-1.12	0	-1.43
Combination 12	4	4	0	2.50	-0.92

Note. All Cohen's d_s's entail effect sizes of the mean number of events compared to the pre-measurement. Means are calculated per research phase.

** $p < 0.05$, * $p < 0.1$.

Table 7

Mean Number of Events in R3 of the SSGs for ASNV and CTA

	Pre-Measurement	Intervention		Post-Measurement	
	M	M	Cohen's d_s	M	Cohen's d_s
Combination 7	7	5.50	-0.42	3.50	-1.15
Combination 8	3	3.25	0.21	2.50	-0.40
Combination 9	2.50	3.25	0.38	1	-0.66
Combination 10	6.50	3.25*	-1.39	1.50	-3.33
Combination 11	3	4.75	0.48	9	4.84
Combination 12	1	2	0.56	5	4

Note. All Cohen's d_s's entail effect sizes of the mean number of events compared to the pre-measurement. Means are calculated per research phase.

** $p < 0.05$, * $p < 0.1$.

Although the MC analyses showed only two mean differences were significant (Table 6 and 7), compared to PRE, in total (POST for combination 7 in R1 and INTER for combination 10 in R3), some of the trajectories do adhere to our expectations. As can be seen in table 6, three out of the six combinations had a higher mean number of events in R1 during the intervention than during the pre-measurement. Two of those were even higher in the post-measurement. Moreover, combination 7 had a higher mean number of events in POST than in the PRE, even though the mean number of events was lower in the INTER than in PRE. To answer our third research question, we did not only hope to see an increase of the mean number of events in R1, but also a decrease of events in R3 in the SSGs for both INTER and POST (Table 7). This was the case for only two teacher-class combinations in the intervention. However, four combinations had a lower mean number of events in R3 in the POST compared to the PRE (Table 7). Moreover, the effect sizes are medium to large for these differences (Lakens, 2013). This gives us some indication that the mean number of events decreases in region 3 for four teacher-class combinations. Combined, we see meaningful coherence for ASNV and CTA for two combinations (7 and 9), slightly meaningful coherence for one combination (10), and not meaningful coherence for three combinations (8, 11, and 12).

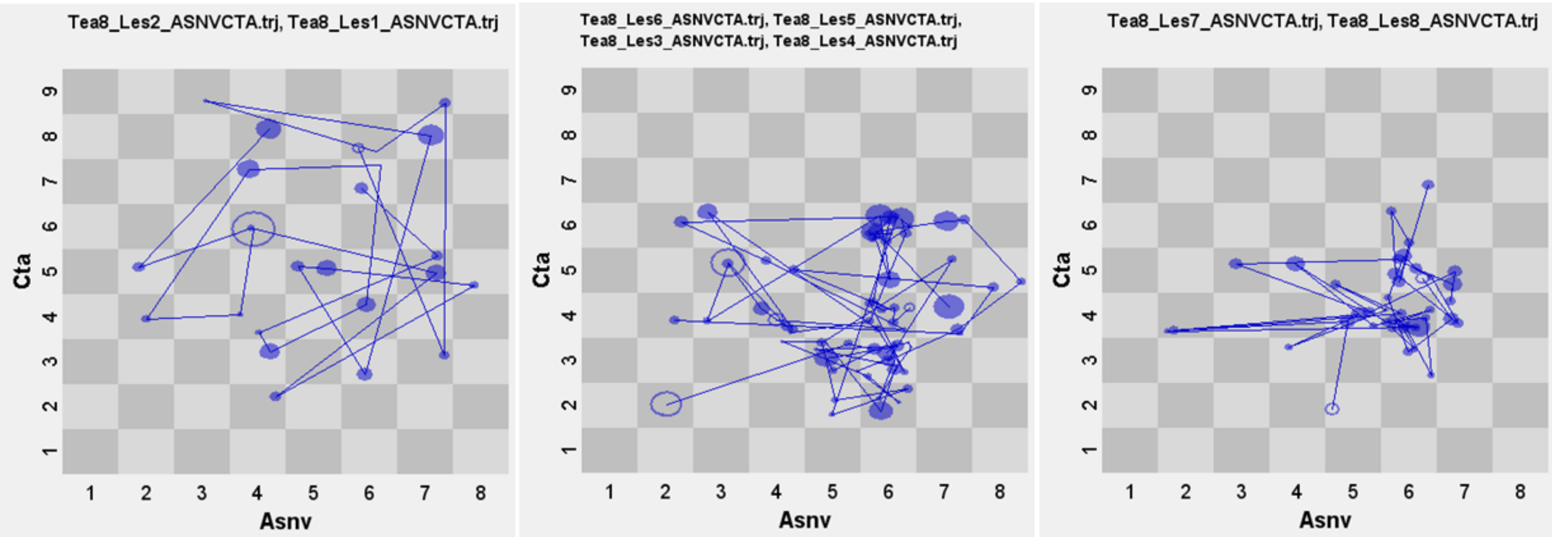
State Space Grids (visual interpretation)

Two representative cases of SSGs that show a trajectory of a teacher-class combination will now be described. The choice for combination 8 is based on the negative correlation (Table 5). Combination 10 was not chosen because of the correlation, but because the scores on ASNV and CTA on INTER and POST were all significant compared to PRE. Other SSGs can be found in Appendix D.

Teacher and class 8. While teacher 8 increased in her average ASNV score, class 8 decreased in their level of CTA throughout the experiment. The SSGs in Figure 5 are in correspondence with this. One can clearly see that the nodes are spread on the SSG grid of the pre-measurement, and then starts to cluster from the intervention period onwards. In the intervention period one can see that the interactions start to move in row of ASNV code 6, and in the post-measurement the nodes are clustered even more. However, the nodes move toward the right down corner, which indicates higher ASNV of the teacher, but relatively low CTA of the corresponding class. This is in contrast of what was expected. We hypothesized that higher ASNV would lead to higher CTA, but there is no sign of that in these SSGs. Therefore, for this teacher-class combination, we would reject the hypothesis.

Figure 5

SSG of Teacher 8

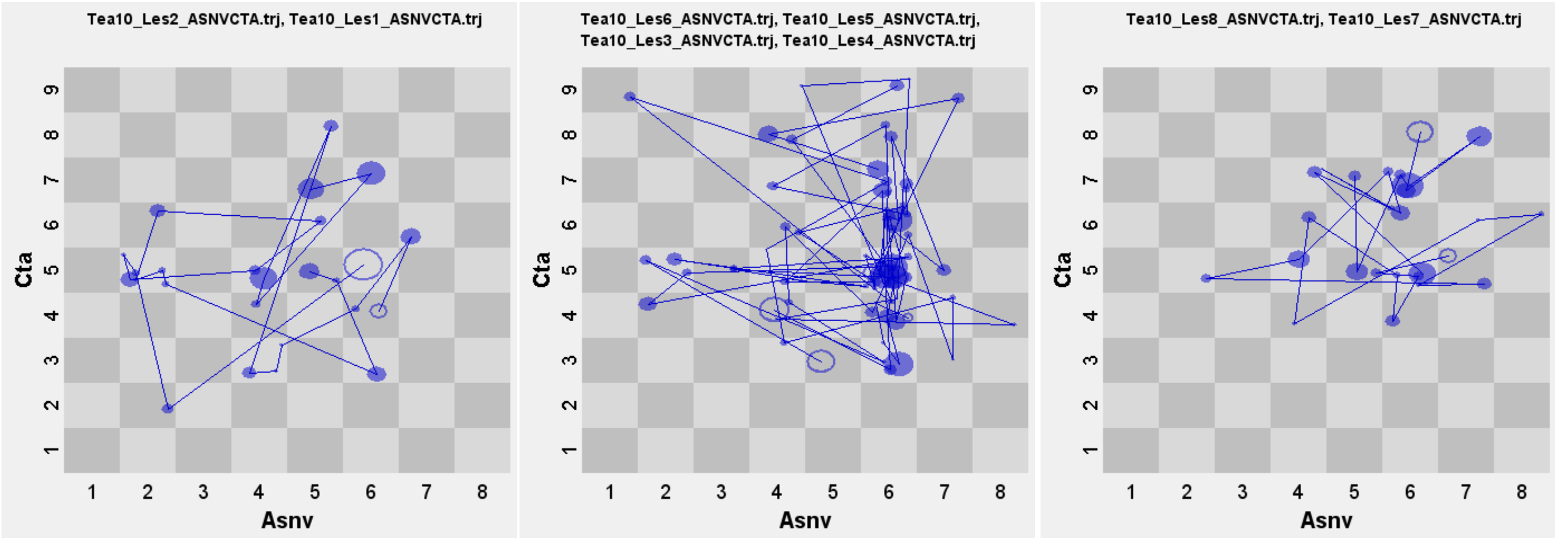


Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Teacher and class 10. This combination of teacher and class is the only one in which both the teacher and the class showed steadily increasing levels of ASNV and CTA, respectively. Therefore, we chose to show the SSGs of this teacher-class combination. In contrast to the SSG of teacher 8 and her class, these SSGs show progression to the upper right corner. During INTER there are a lot of codes in the ASNV row 6. When we look at the nodes in POST, we see interaction pairs in ASNV 6 and 7 with CTA 7 and 8, something we did not see in PRE. For this teacher-class combination there does seem to be a positive coherence between ASNV and CTA. This is something that also appeared in the mean number of events (Table 7 and 8). The MC analysis showed that the difference between PRE and INTER was significant for the decrease of events in region 3. However, because of the weak correlation (ρ = .31), we can only partly accept the hypothesis for this combination. Over time, both ASNV and CTA seem to progress towards higher levels, but not always in strong correlation with each other. In contrast, we did see more positive interactions between class and teacher.

Figure 6

SSG of Teacher 10



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right.

Summarized results on ASNV and CTA coherence

Of the teachers that had increases in ASNV over all research phases, only teacher 10 had slightly meaningful coherent interactions with her classes, with simultaneously increasing levels of CTA in her class on a group level. Teachers 7 and 9 had meaningful coherent interactions with their classes that increased over time, but on a group level we did not see higher CTA scores in all research phases. Only increases were seen in POST. In contrast, teachers that had increases in ASNV in all research phases (8 and 12), but had no higher CTA scores in their classes, had no increase in meaningful coherent interactions either. More so, teacher and class 11 both had increases during INTER in ASNV and CTA, respectively, but showed no sign of increasing coherence between the two. There seems to be some coherence between ASNV and CTA, but not always when both show increases over time. Therefore, the combination of correlations and mean number of events provide, slight support for the hypothesis

Conclusion & discussion

In this study we first looked at the development of teachers after a coaching intervention aimed at improving non-verbal autonomy support (ASNV). Our hypothesis was: 'Expressed non-verbal autonomy support by the teacher can be increased through coaching'. We compared both INTER and POST with PRE to determine if there was progression and if this progression lasted. The increasing percentage of higher (6, 7, and 8) and decreasing percentages of lower (1 and 2) ANSV codes during INTER and POST, were a first sign in support of this hypothesis. In addition, the MC analyses and Cohen's d_s's showed meaningful increases in ASNV. Subsequently, the positive slope of the MMs of five teachers, and the conclusion that the effects of coaching on ASNV were meaningful, as derived from the mean differences, all lead us to accept this hypothesis. Therefore, we can conclude that teachers are very well able to be taught to show more non-verbal autonomy support. Something that has been priorly stated by Reeve and Cheon (2021) as well, but had not been established in music education (Simones, 2017).

That four of the five teachers that had significantly improved on ASNV during INTER, remained significantly improved during POST, shows signs of durability of the coaching effect. Only teacher 11 did not maintain their acquired increase in ASNV. Due to the Covid-19 pandemic that emerged during assessment of the videos, teacher 11 had only one post-measurement, making it more susceptible for outliers. The loss of retention could be a skewed image representation of the real level of that teacher during that time. However, in general, the pandemic could explain some of the results, but it also offers insight in the sustainability of the coaching effect. In contrast, the pandemic impacted teachers 10, 11, and 12 during this study (recordings of teacher 7, 8, and 9 just finished). For these teachers, coping capabilities to the stress of a pandemic may also be a factor, as well as having to deal with rigorous changes in work environment and life in general.

Furthermore, little experience in both teaching and music could possibly have caused a slower learning trajectory. For example, teacher 7 was a relatively young and less experienced teacher, and did not have a musical background. She did not show significantly higher ASNV scores during the INTER, but did, however, during POST. Coachability of a teacher might therefore be dependent on the (musical) experience of that teacher. More qualitative analysis is needed to make this conclusion, as this could provide more insight in interindividual uptake of training material.

Our second hypothesis was that over the course of the research, we would see an increase of convergent thinking and acting (CTA) in music lessons for primary school children. Only for POST we began to see an upwards shift in the percentual distribution for CTA compared with PRE. For some of the classes we could distinguish a progression, but for most we could not. This was further emphasized by the MC analyses and Cohen's d_s's. These showed us that, on group level, the results for CTA were just slightly meaningful for POST, but not meaningful for INTER. That is, only three classes had higher average CTA scores (of which two were significant) during INTER compared to PRE, and two of those remained higher than PRE during POST (only one significant). Furthermore, the frequencies gave some indication of growth in CTA, but the visualization did not convincingly. Despite the slightly meaningful results for POST, the lack of meaningfulness of the results for INTER, in combination with lack of further support of the frequencies, the visualization, and the significance of mean differences, we reject our hypothesis.

An explanation for the mixed results for CTA could be that it might have taken some time for the children to get used to the music methods used in the intervention. That is, the children might have had difficulties with the differences between the lesson type during PRE (in which teachers were asked to teach as usual) and INTER and POST (which were predominantly supportive of creativity), having impact on the relative ease of achieving higher CTA scores. During lessons in which the emphasis was on singing, it appeared to be easier to deal with more complex musical patterns as compared to creative musical tasks. This might have resulted in higher CTA-levels during PRE. Another explanation might be that, with more emphasis on creation than imitation, music tasks were getting increasingly more difficult during lessons. On the one hand, this gave the children the opportunity to show higher levels of CTA, but on the other hand the tasks might have become too difficult for some of the children, taking place outside their zone of proximal development (e.g., McLeod, 2019), therefore, hindering their performance. This might even be dependent on the age and musical experience of a child. However, regardless of these sidenotes, the CTA scores were slightly meaningful for POST while they were not for INTER, therefore, it would be interesting to see what would happen if the research was extended past the 8 lessons that it comprised in the current study.

Finally, we hypothesized that we would see a positive correlation between increases in ASNV and CTA. Results on the Spearman’s rho correlations in combination with the mean number events of the SSGs, showed that there was meaningful coherence between ASNV and CTA for teacher-class combination 7 and 9, and slightly meaningful coherence for combination 10. The others were not meaningful. However, on group level, (slightly) meaningful coherence was not always accompanied with increases in CTA in the class. This leads us to partly accept the hypothesis. To some extent, we do seem to be able to stimulate positive interactions between teacher ASNV and class CTA through an intervention, but it is not said that increases in ASNV always lead to increases in CTA.

Progression in the teacher on ASNV does not necessarily mean improvement in CTA for the class and that the relation between the two can also be negative, as is shown by the correlation and SSG of teacher-class combination 8. It might be that this teacher focused too much on ASNV and thereby paid less attention to the other elements of teaching. Another explanation could be that ASNV is weaker correlated to CTA than to divergent thinking and acting (DTA). That is, difficult rhythmic patterns may ask more direct support by the teacher instead of, for example, an expectational gaze. In case of the latter, scaffolding might have been insufficiently applied by the teachers, or the teacher had to switch a lot between the amount of support, exemplifying the possibly ambiguous relation between ASNV and CTA.

Noticeably, when we look at the coherence between ASNV and CTA, the only class to have significantly higher average CTA scores during INTER *and* POST belonged to the teacher that was among the two lowest ASNV scoring teachers during PRE (teacher-class combination 10). Visual inspection of the SSGs gives the impression of a link between increasing CTA scores in children and the great amount of room for improvement in the teacher in some cases. However, the correlation of ASNV and CTA, in this case, was weak (ρ = .31). This might indicate that, although both teacher and class progressed, it might not always have been the direct interaction between the two that lead to improvement. The delayed POST may have led, in some but not all cases, to a degeneration of results.

To conclude on the growth of ASNV and CTA, and the coherence between the two: What we noticed is that we could not accept the hypothesis for RQ2, while we did find significant increases for ASNV. A possible explanation could be the impact of other variables. As the behavior of children, and human behavior in general, is dependent on a dynamic interaction of components (Fischer & Bidell, 2006), there might be other variables that might have driven CTA in children in another direction than the teacher ASNV. For example, we looked at the non-verbal autonomy support of the teacher, but it is not said that non-verbal and verbal autonomy support are synchronous. Overall, it might be too precarious to directly link the impact of teacher ASNV on CTA in children.

Limitations & future research

When we look at the circumstances of the study, a first limitation concerning the data, is that all teachers in this study were female, children might react different on male teachers. This might even differ according to gender of the child. Although there have been no signs of the influence of a standalone gender effect for teachers on the general performance of children (Sabbe & Aelterman, 2007), and performances did not seem to increase for boys who were exposed to more male teachers (De Zeeuw et al., 2014), we do not know if this is also the case for (creativity in) music. Therefore, results from this study should be taken carefully when generalizing to male teachers. Moreover, Sabbe and Aelterman (2007) debate that gender attributes to a teacher’s identity and that this is also socially and culturally dependent. Hence, another important question to ask ourselves is whether culture affects teacher training in autonomy support as well. This study is conducted in the Netherlands, and in western cultures, which are more individualistic, children may have more confidence in showing autonomy and to question their teachers. While in non-western cultures, were there is often a stronger hierarchical foundation, it is less acceptable for children to express themselves without the consent of an adult or a teacher. Therefore, autonomy supportive teaching might be more beneficial in some cultures than in others. On top of that, children in the Netherlands might react differently on this intervention based on their cultural background. The extent to which education fits the cultural background of a child can be important for its academic performance, and therein, culturally responsive teaching is an important factor (Gay, 2002). This asks for caution when translating this research to all kinds of cultures. Future research can extend to this by differentiating the effects of ASNV on musical expression in children from varying cultural backgrounds, and by looking at other socially culturally influences on teacher behavior.

A third limitation is that the researcher who did the intervention was also present during the recording of the videos. This might have led to a priming effect. This could be troublesome because we do not know if the teachers would have shown the same behavior if the researcher was not present. The presence of the researcher could have been a subconscious reminder for the teachers in an otherwise ‘natural’ environment. Moreover, the teacher and researcher worked together during the intervention. This might have created a perception of each other that could be both positive and negative. There is always a jeopardy for some influence of the researcher or response biases, but in this case possibly even stronger because of the intensive collaboration that occurred in advance. The influence could be both direct and indirect. In a direct way during the recording of the videos and in an indirect way in the sense that the affect of the teacher could influence her motivation to put effort in. The prior could be prevented by having the video recordings made by an independent person which the teacher did not interact with yet. Assessment of the satisfaction with the collaboration during the study of each teacher, might also provide a possible solution and could provide more insight in the results.

More than just the circumstances during the research, the dataset and analyses used, also have some limitations. The use of averages for ASNV and CTA gave information about the levels of the teachers and children. On top of that, the analyses through the use of MMs and SSGs provided additional in-depth information about the developmental trajectories of the variables. However, apart from their practical use, these analyses encompass some limitations one should be aware of as well. For example, using averages for ASNV and CTA and focusing on the research phases, tells us something about the general level in each research phase, and the MMs only added insight on the height of ASNV to that. They do not provide a complete picture of the growth of the teachers and classes, as growth can occur on multiple aspects. The frequencies provide some insight in the occurrence of lower scores in each research phase, and the SSGs does provide information about interactions including the lower scores, but both lack more detailed information about the trajectories of lower scores. A possible analysis that could provide more information in the future, is not only looking at the moving maximum, but also at a moving minimum and using this bandwidth to visually interpret progression per lesson. This could give detailed information not currently provided by this study.

Moreover, some detailed data is lost while preparing data for the SSGs. The SSGs provide unique insight in the interaction patterns of teacher and students, but to do so for two separate sets of codes, sometimes causes several ASNV codes to be linked to fewer CTA codes, and vice versa. That is, by connecting the highest perceived ASNV level within a turn with the subsequent CTA level, potentially relevant information of the performance of the child is lost. Therefore, it is difficult to make assumptions on basis of the SSGs about the direct effect of ASNV on CTA. However, the SSG showed to be useful to make assumptions about general interaction patterns in a certain timeframe.

Another example of a limitation of the data in this study, is that of the small amount of lessons. For the Cohen's d_s for the mean number of events this meant that there were few lessons from which to take information. For example, the pre-measurement had only two lessons on which the mean number of events was calculated. Because of the small *n*, effect sizes can become large very quickly. Moreover, the small number of lessons also affect the sensitivity for outliers. For example, class 8 showed relatively stable scores on CTA until lesson 6. At lesson 6 they scored a lot lower than in the other lessons. Even though the time-serial datapoints and the analyses used provide enough data for the individual cases and take the small samples in regard, such an outlier in a small sample of 8 lessons could have a big impact, when comparing based on averages. Therefore, it might be interesting to have qualitative analyses of outliers like these lessons. A small sample of teachers, like in this research, provides better or additional opportunity for qualitative analysis (e.g., Kupers & Van Dijk, 2020). That is, qualitative research could provide more in-depth information about what exactly happens in a primary school classroom. Focusing on a small number of children instead of a whole class could facilitate more detailed research.

In this study we focused on the impact of ASNV on CTA to study the coherence between the two. However, several variables that might influence children CTA, which are not included in this study, are: (1) verbal autonomy support, as verbal autonomy support may be leading and non-verbal autonomy support may be limited to additional value to verbal autonomy support (Novack & Goldin-Meadow, 2015); (2) teacher and student affect, as a more cheery mood might work stimulating, and a more steep mood might work obstructing for the supporting of autonomy (in case of the teacher), and the susceptibility of receiving autonomy (in case of the child); (3) prior musical experience in teacher and children, as musical experience might elicit confidence to explore for the child, and confidence to teach autonomously for the teacher.

Finally, during the assessment of the research in 2020, the Covid-19 pandemic caused for disturbances in the continuation of this research. Minimal physical contact was advised by the Dutch government, and this also accounted for the schools. One of the effects that it had on this research was that no video recording was made of the final lesson of teacher 11. It also caused variability in the time between the recordings of the post-measurement for some other teachers.

This research has shown that teachers can be coached in using ASNV, and that there are signs that interactions in ASNV and CTA tend to move together after an intervention.

Future research is needed to better understand the factors influencing the direction. Doing so, would help music education to guide teacher-student interaction towards positive outcomes.

References

Alves-Oliveira, P., Arriaga, P., Xavier, C., Hoffman, G., & Paiva, A. (2021). Creativity Landscapes: Systematic Review Spanning 70 Years of Creativity Interventions for Children. *The Journal of Creative Behavior*. https://doi.org/10.1002/jocb.514

Aróstegui Plaza, J. L. (2016). Exploring the global decline of music education. *Arts Education Policy Review, 117*(2), 96–103.

Benthem, J. van, Dijkgraaf, R., & De Lange, J. (2005). TalentenKracht [Talent Empowerment]. Utrecht, Freudenthal Institute.

Bishop, L. (2018). Collaborative Musical Creativity: How Ensembles Coordinate Spontaneity. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01285

Bos, J. and Steenbeek, H., *Mediacoder, A Simple Application For Coding Behavior within Media Files*. University of Groningen, Groningen, The Netherlands, 2010.

Castillo, A., Domanick, A., & Matos, M. (2014, March 14). 50 Most Important People in EDM. Rolling Stone. Retrieved 24 April 2021, from https://www.rollingstone.com/music/music- lists/50-most-important-people-in-edm-30822/skrillex-5-49178/

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi-org.proxy-ub.rug.nl/10.1207/S15327965PLI1104_01

De Dreu, C. K. W., Baas, M., & Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: Toward a dual pathway to creativity model. *Journal of Personality and Social Psychology, 94*(5), 739–756. https://doi.org/10.1037/0022-3514.94.5.739

De Zeeuw, E. L., van Beijsterveldt, C. E., Glasner, T. J., Bartels, M., de Geus, E. J., & Boomsma, D. I. (2014). Do children perform and behave better at school when taught by same-gender teachers? *Learning and Individual Differences*, *36*, 152–156. https://doi.org/10.1016/j.lindif.2014.10.017

Evolutie van elektronische muziek - Deel 1/3 [Evolution of electronic music - Part 1/3]. (2020, August 19). Retrieved 24 April 2021, from https://www.thomann.de/blog/nl/evolutie-van-elektronische-muziek-deel-1-3/

Evolutie van elektronische muziek - Deel 2/3 [Evolution of electronic music - Part 2/3]. (2020, August 19). Retrieved 24 April 2021, from https://www.thomann.de/blog/nl/evolutie-van-elektronische-muziek-deel-2-3/

Evolutie van elektronische muziek - Deel 3/3 [Evolution of electronic music - Part 3/3]. (2020, August 20). Retrieved 24 April 2021, from https://www.thomann.de/blog/nl/evolutie-van-elektronische-muziek-deel-3-3/

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology* (6th ed., Vol. 1, pp. 313-399). New York: Wiley.

Gay, G. (2002). Preparing for Culturally Responsive Teaching. *Journal of Teacher Education, 53*(2), 106–116. https://doi.org/10.1177/0022487102053002003

Harthan, J. (1943). Eighteenth-Century Flute Music. *Music & Letters, 24*(1), 35-42. Retrieved April 24, 2021, from http://www.jstor.org/stable/728609

Hendriks, L. H., Steenbeek, H. W., & Bisschop Boele, E. H. (2018, Mar 29). Curious Minds Muziekeducatie: Muzikale creativiteit in de klas. Werkboek voor leerkrachten in de middenbouw van het basisonderwijs.

Juslin, P. N. (2003). Five Facets of Musical Expression: A Psychologist’s Perspective on Music Performance. *Psychology of Music*, *31*(3), 273–302. https://doi.org/10.1177/03057356030313003

Kondo, S., & Wiggins, J. (2018). Learner Agency in Musical Creative Process and Learning. *Creativity in Music Education*, 17–33. https://doi.org/10.1007/978-981-13-2749-0_2

Kupers, E., & van Dijk, M. (2020). Creativity in interaction: the dynamics of teacher-student interactions during a musical composition task. *Thinking Skills and Creativity*, *36*, 100648. https://doi.org/10.1016/j.tsc.2020.100648

Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2018). Children's Creativity: A Theoretical Framework and Systematic Review. *Review of Educational Research*, *89*(1), 93–124. https://doi.org/10.3102/0034654318815707

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

La Guardia, J. G., & Ryan, R. M. (2007). Why identities fluctuate: Variability in traits as a function of situational variations in autonomy support. *Journal of Personality*, *75*(6), 1205–1228. https://doi-org.proxy-ub.rug.nl/10.1111/j.1467-6494.2007.00473.x

Malinin, L. H. (2019). How Radical Is Embodied Creativity? Implications of 4E Approaches for Creativity Research and Teaching. *Frontiers in Psychology*, *10*, 1–12. https://doi.org/10.3389/fpsyg.2019.02372

Marshall, R. (2005). *18th-Century Keyboard Music (Routledge Studies in Musical Genre)* [EPub] (2nd ed.). Retrieved from http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=e000xww&AN=598533&site=ehost-live&scope=site&ebv=EK&ppid=Page-__-2

Mcleod, S. (2019, March 24). *The Zone of Proximal Development and Scaffolding*. Simply Psychology. https://www.simplypsychology.org/Zone-of-Proximal-Development.html#:%7E:text=%22the%20distance%20between%20the%20actual,86)

Mieras, M. (n.d.). Wat onderzoek ons leert over muziek op school [What research teaches us about music at school]. Meer Muziek in de Klas. Retrieved 4 July 2021, from https://www.meermuziekindeklas.nl/uploads/pdf/MMiK_-Mark-Mieras_DEF.pdf

Montagu, J. (2017). How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, from Its Earliest Stages. *Frontiers in Sociology, 2,* 1–12. https://doi.org/10.3389/fsoc.2017.00008

Numeracy, Maths and Statistics - Academic Skills Kit. (n.d.). Retrieved 8 January 2022, from https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/strength-of-correlation.html

Novack, M., & Goldin-Meadow, S. (2015). Learning from Gesture: How Our Hands Change Our Minds. *Educational Psychology Review*, *27*(3), 405–412. https://doi.org/10.1007/s10648-015-9325-3

Puth, M. T., Neuhäuser, M., & Ruxton, G. D. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, *102*, 77–84. https://doi.org/10.1016/j.anbehav.2015.01.010

Reeve, J., & Cheon, S. H. (2021). Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice. *Educational Psychologist*, *56*(1), 54–77. https://doi.org/10.1080/00461520.2020.1862657

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

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*, 141–166. https://doi-org.proxy-ub.rug.nl/10.1146/annurev.psych.52.1.141

Sabbe, E., & Aelterman, A. (2007). Gender in teaching: a literature review. *Teachers and Teaching*, *13*(5), 521–538. https://doi.org/10.1080/13540600701561729

Simon, P. D., & Salanga, M. G. C. (2021). Validation of the Five-item Learning Climate Questionnaire as a measure of teacher autonomy support in the classroom. *Psychology in the Schools*. Published. https://doi.org/10.1002/pits.22546

Simones, L. L. (2017). A framework for studying teachers' hand gestures in instrumental and vocal music contexts. *Musicae Scientiae*, *23*(2), 231–249. https://doi.org/10.1177/1029864917743089

Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting Autonomy in the Classroom: Ways Teachers Encourage Student Decision Making and Ownership. *Educational Psychologist*, *39*(2), 97–110. https://doi.org/10.1207/s15326985ep3902_2

Turner, J. C., & Christensen, A. L. (2020). Using state space grids to analyze teacher–student interaction over time. *Educational Psychologist*, *55*(4), 256–266. https://doi.org/10.1080/00461520.2020.1793763

Todman, J.B., & Dugard, P. (2001). *Single-case and small-n experimental designs: A practical guide to randomization tests*. Mahwah (NJ): Erlbaum.

Van Geert, P. (2008). Nonlinear Complex Dynamical Systems in Developmental Psychology. *Chaos and Complexity in Psychology*, 242–281. https://doi.org/10.1017/cbo9781139058544.009

Van Vondel, S., Steenbeek, H., van Dijk, M., & van Geert, P. (2017). Ask, don't tell; A complex dynamic systems approach to improving science education by focusing on the co-construction of scientific understanding. *Teaching and Teacher Education*, *63*, 243–253. https://doi.org/10.1016/j.tate.2016.12.012

Viera A.J., Garrett J.M. (2005, May). Understanding interobserver agreement: the kappa statistic. *Fam Med*, *37*(5), 360-3. PMID: 15883903.

Webster, P. R. (1990). Creativity as Creative Thinking. *Music Educators Journal*, *76*(9), 22–28. https://doi.org/10.2307/3401073

Webster, P. R. (2002). Creativity and music education: Creative thinking in music: Advancing a model. *Creativity and music education*, *1*, 33.

Webster, P. R. (2009). Children as creative thinkers in music. *The Oxford handbook of music psychology*, 421-428.

Yusof, F. B., & Tan, A. G. (2018). Teaching Music in the Early Childhood Classroom for Convergent Creativity: Views from a Meta-synthesis. *Creativity in Music Education*, 245–266. https://doi.org/10.1007/978-981-13-2749-0_18

Appendix A

Global coding scheme of non-verbal autonomy support by the teacher

Level	Code	Description	Example
Low autonomy support	1: Stop sign	Disrupting/stopping students: The teacher wants silence The teacher disapproves of how something is played/done Note: Not when a play or turn has ended	Taking an instrument from a student Giving a stop sign (by raising their hand)
	2: Modeling & instructing	Literally showing how something needs to be done. Giving an example of how something must be played/sung according to the teacher or the task instructions. Students observe and listen or play along with the audiovisual example. Aimed at eliciting imitation. Students follow the teacher	Showing how a something is played Rhythmically prompting words by the teacher Instructive gesture. For example, pointing at where a drum must be hit
Medium Autonomy support	3: Participative support	Offering audiovisual support, at which students offer their own ideas. Teacher plays/sings along, but involves students based on their own ideas	Clapping along with the rhythm of the student
	4: Representational gesturing	Gestures to express (musical) concepts of tasks during verbal explanation	Moving hand up and down to express tone height
Medium-High Autonomy support	5: General non-verbal activity and movement	Slightly stimulating activity, that not (yet) explicitly stimulates musical creativity, like moving, dancing or rhythmically shaking their head along with the music	Teacher moves rhythmically along with the music from the speakers Inviting a student to the front of the class
	6: Thought eliciting gestures and conducting (during student play)	Offering visual information in advance of performing a musical task by the students by visualizing the (musical) targets. Students need to	Clapping the beat Hand/arm movements for volume, tempo, and tone height

		interpret the visual information themselves and translate it to action	Giving students a turn by pointing at them
High autonomy support	7: Observing and offering space to act or create	Teacher offers students space to (continue) play or exploration on basis of the ideas of students own ideas and the teacher follows students attentively and explorative by listening and observing actively and sometimes give non-verbal coaching cues. Minimal guidance in gesture, mimic, or movement. Autonomous creative musical behavior eliciting support	Actively listening to the students play without interrupting Showing involvement in facial expression Moving, dancing, or nodding along with student play
	8: Encouragement: high stimulation of student play in gesture, movement and/or facial expression	Encouraging by stimulating behavior in gesture, movement and/or facial expression. Aimed at eliciting explorative behavior of the students towards musical ideas and taking risks in the process	- Inviting gestures - Thumbs up - Positive and/or facial expression full of expectation

Appendix B

Global coding scheme of convergent thinking/acting by children in music lessons

Tier	Code	Description	Verbal indicators
Sensory Motor Actions, explorative behavior aimed at making sounds. Irregular, explorative and focused at making sound	1: Single Sensory Motor Actions:	The child looks at how it can make sound with the instrument and/or makes a single sound. The child acts with the instrument, looks at it, turns it around, looks at it from multiple angles, and seems to consider taking musical action. This could be prior to musical action, but also in between two actions.	Simple observations: - "Hey!" - "That sounds strange/nice/dim."
	2: Sensory Motor Actions Mapping	The child makes purely explorative sounds with the instrument. Musical play is irregular. It can be hesitant, but also quick and shallow. Before, after, or in between, verbal reflection can take place in which differences of sounds are noticed. There is no coherence between sounds; no pattern to hear.	Comparisons of sounds: - "It is also possible this way." - "This way it sounds nicer."
	3: Sensory Motor Actions Systems	The child shows understanding of a sound effect of (a) specific action(s) with the instrument by performing the action intentionally. The link of an action and its consequence of a specific action with a specific sound, is noticed. The child knows how to make a	Understanding of a sound effect. For example, the child can give an explanation. Sometimes, prior to an action, an intention or a choice is given. But there are no signs yet of rhythm or regularity, emphasis is on making a single sound:

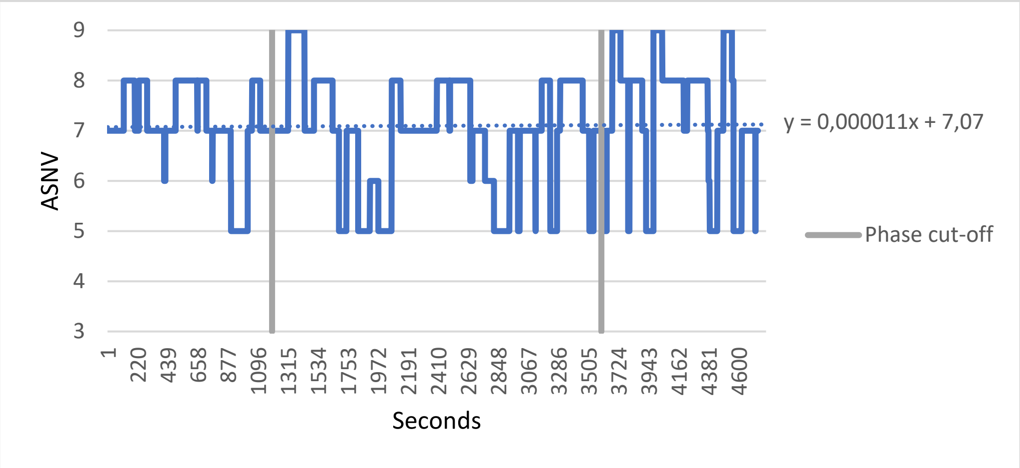
		targeted sound. Sometimes, this can be noticed because a certain action is repeated within the play.	- “Pulling the string and then hitting it.” (with a triangle) - “Hitting your cheek is also possible.” (with body percussion)
Representations, mainly irregular	4: Single Representations: playing a single pattern	First signs of patterns appear. The children combine single sounds and carefully show patterns. Decision criterium: does it sounds a pattern and not as coincidentally combined single sounds?	In between musical action, children can talk or hum along. Often there is a combination of talking and acting musically.
	5: Representations Mapping	Exploration and/or practice of patterns by variations in patterns or exploration of several different patterns. These patterns are expressed as single patterns, not as a coherent whole. There is no sign of a longer musical sentence yet. Variance can be made through varying tone duration, tempo, timbre, tone height or volume (based on the same rhythmic pattern). Sophistication is also exploration of variations. Playing of a not yet fully accomplished performance of a pattern is also exploration of variations. In between play, children can pause and talk/discuss or hum along.	Children can support their own play with text: - “Stomp, stomp, clap.” They can also give explanations of their intentions, compare their different performances, or explain what went wrong. One can often speak of self-correction and/or explanation of why one variation of sounds is better than the other: - “No, this one need to be different.” - “Not like this, we’ll do it again.” - “It is prettier this way.”

	6: Representations Systems	Children can express themselves musically in a single pattern without interruption or stopping. It sounds like a musical whole and is played rhythmically correct and convincing. Children cohere to their own choices in tone duration, timbre, tempo and volume or tone height. The pattern is practiced at least one time before and is played as a coherent whole.	Giving explanation about the order of a musical pattern: - “First stomp with the left and then with the right, followed by a clap.” - “Now everything all at once”
Abstractions, more complex musical concepts originate. Increasingly getting more consistent, repeating of clearly recognizable patterns. Children can create an increasingly complex sound world	**7**: Single Abstractions	First appearances of rhythmic phrases. Children can play a combination of two or more rhythmic phrases at least once.	
	8: Abstractions Mapping	Exploration of variations of musical sentences in which two or more rhythmic patterns are combined. The sentences get longer through repetition.	
	9: Abstraction Systems	Several different musical concepts are combined, which leads to a longer musical sentence. This sentence forms a coherent musical whole and is played convincingly without verbal expressions in between.	

Appendix C

Figure C1

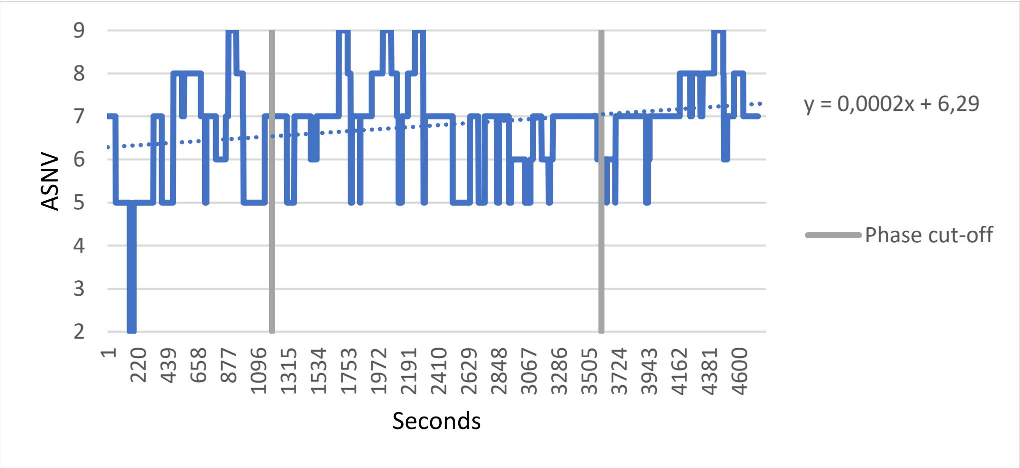
Moving Maximum Teacher 7



Note. This figure contains the moving maximum of teacher 7. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.00001x + 7.07$, with x in seconds.

Figure C2

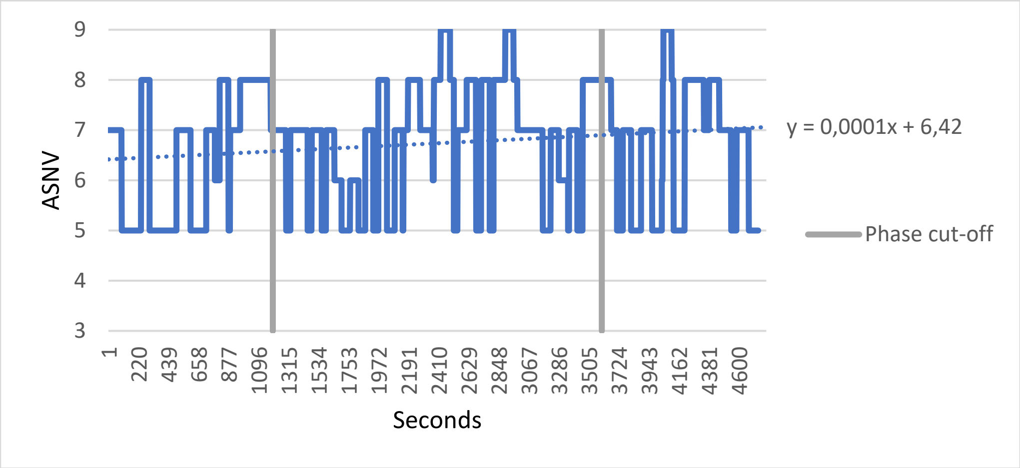
Moving Maximum Teacher 8



Note. This figure contains the moving maximum of teacher 8. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.00002x + 6.29$, with x in seconds.

Figure C3

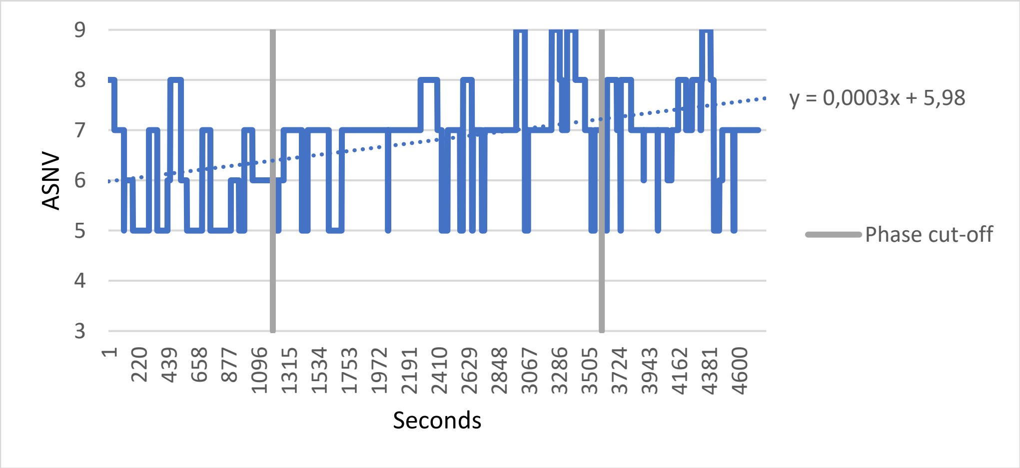
Moving Maximum Teacher 9



Note. This figure contains the moving maximum of teacher 9. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.0001x + 6.42$, with x in seconds.

Figure C4

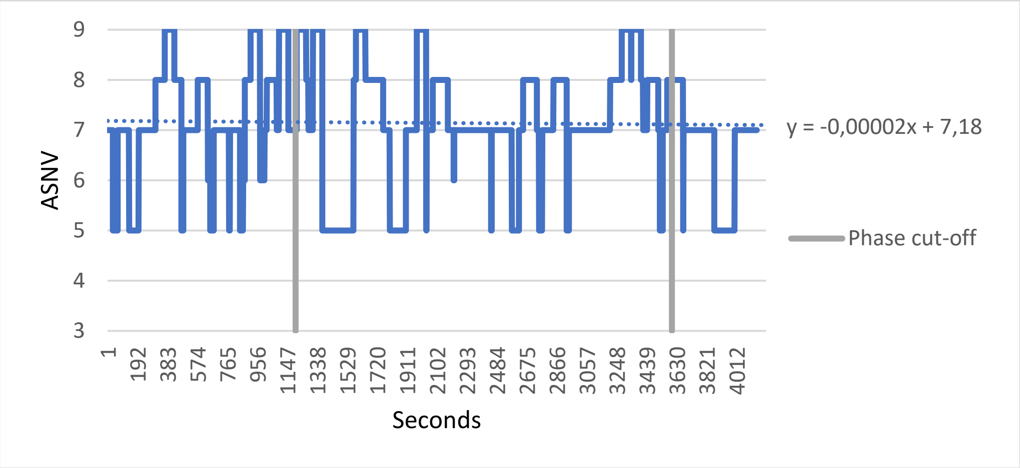
Moving Maximum Teacher 10



Note. This figure contains the moving maximum of teacher 10. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately $y = 0.0003x + 5.98$, with x in seconds.

Figure C5

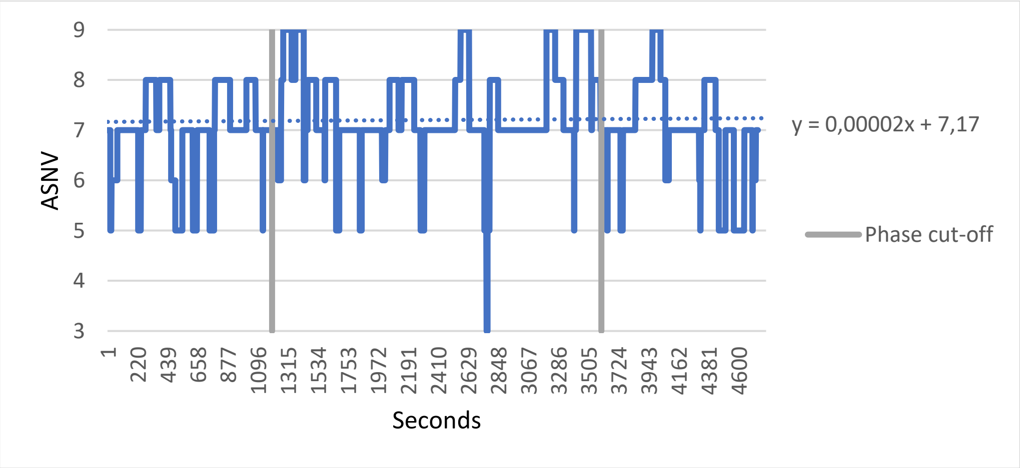
Moving Maximum Teacher 11



Note. This figure contains the moving maximum of teacher 11. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately y = -0.00002x + 7.18, with x in seconds.

Figure C6

Moving Maximum Teacher 12

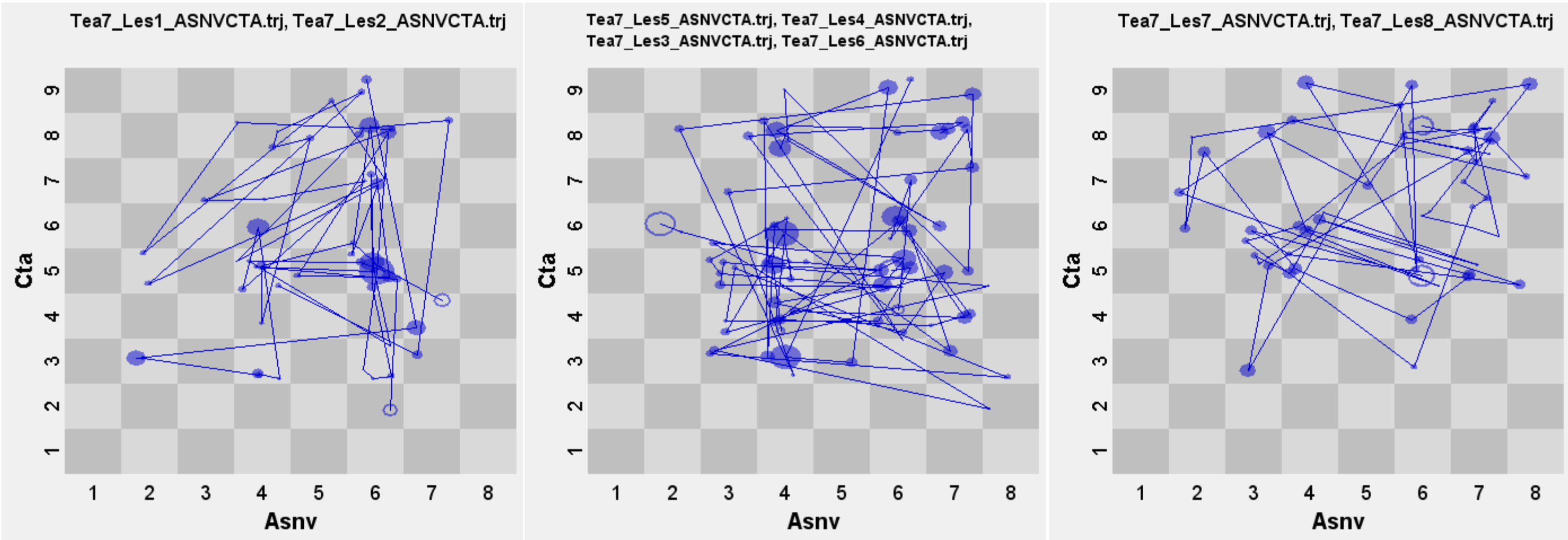


Note. This figure contains the moving maximum of teacher 12. The moving maximum has a possible range of 1 to 9. Research phases are indicated through the grey line. The formula belonging to the trendline is approximately y = 0.00002x + 7.17, with x in seconds.

Appendix D

Figure D1

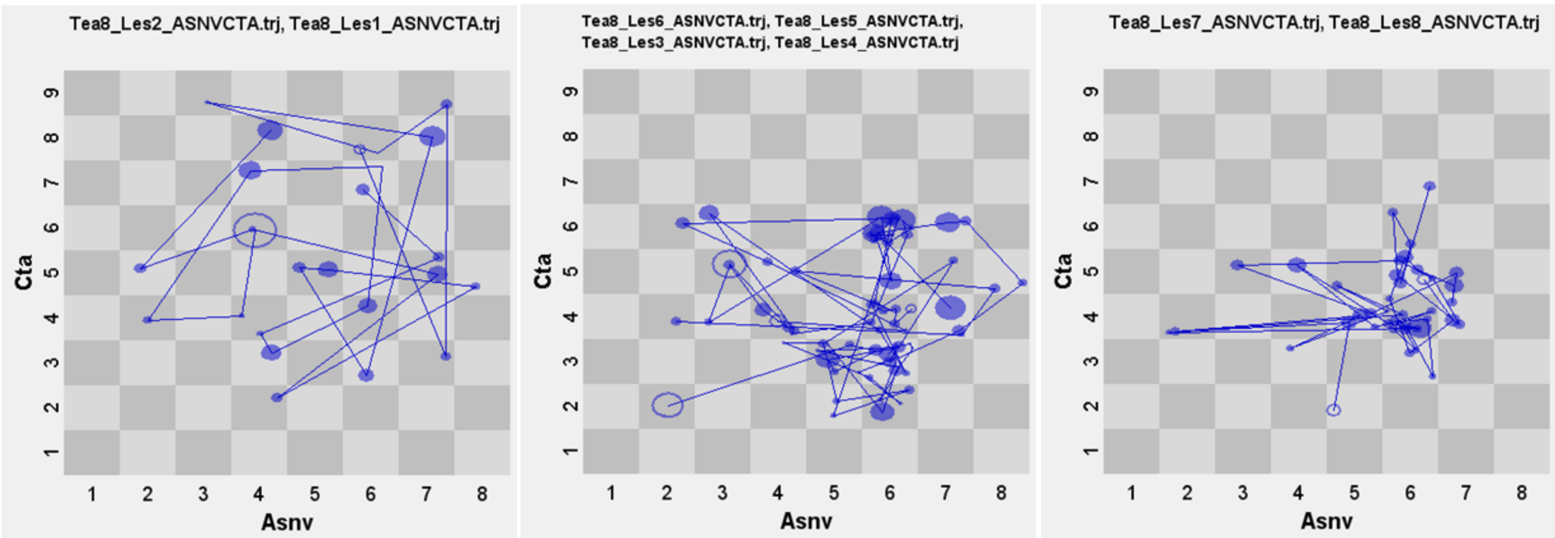
SSG of Teacher 7



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Figure D2

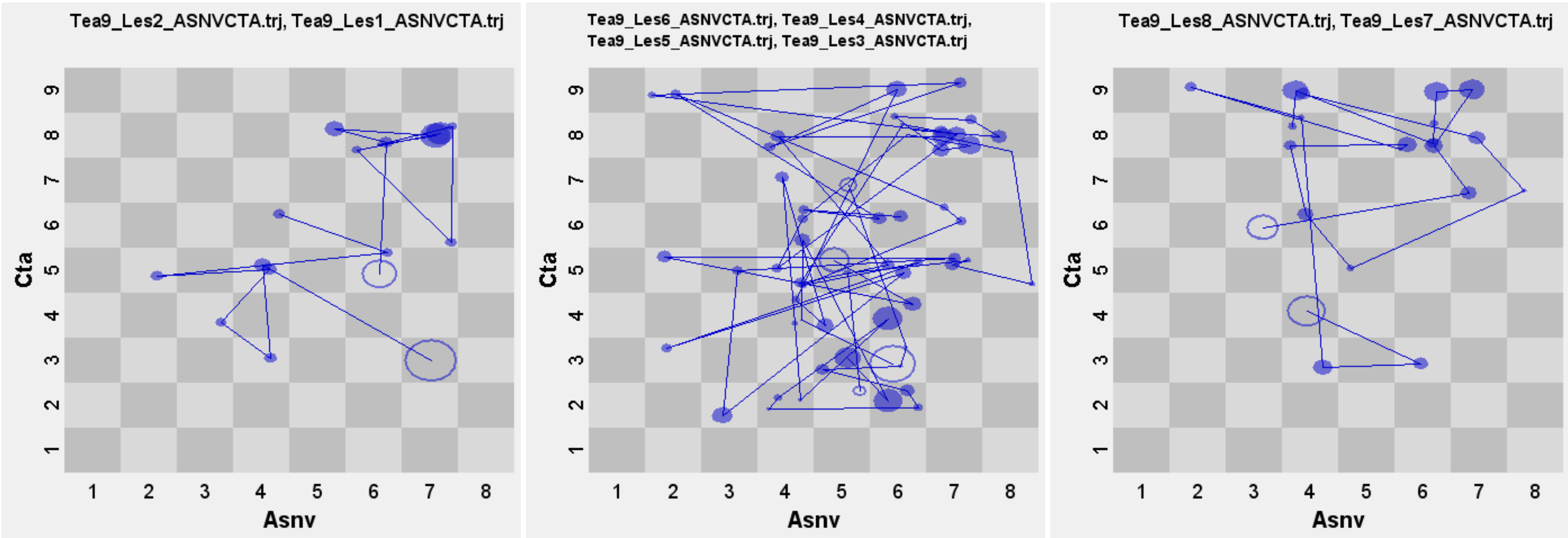
SSG of Teacher 8



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Figure D3

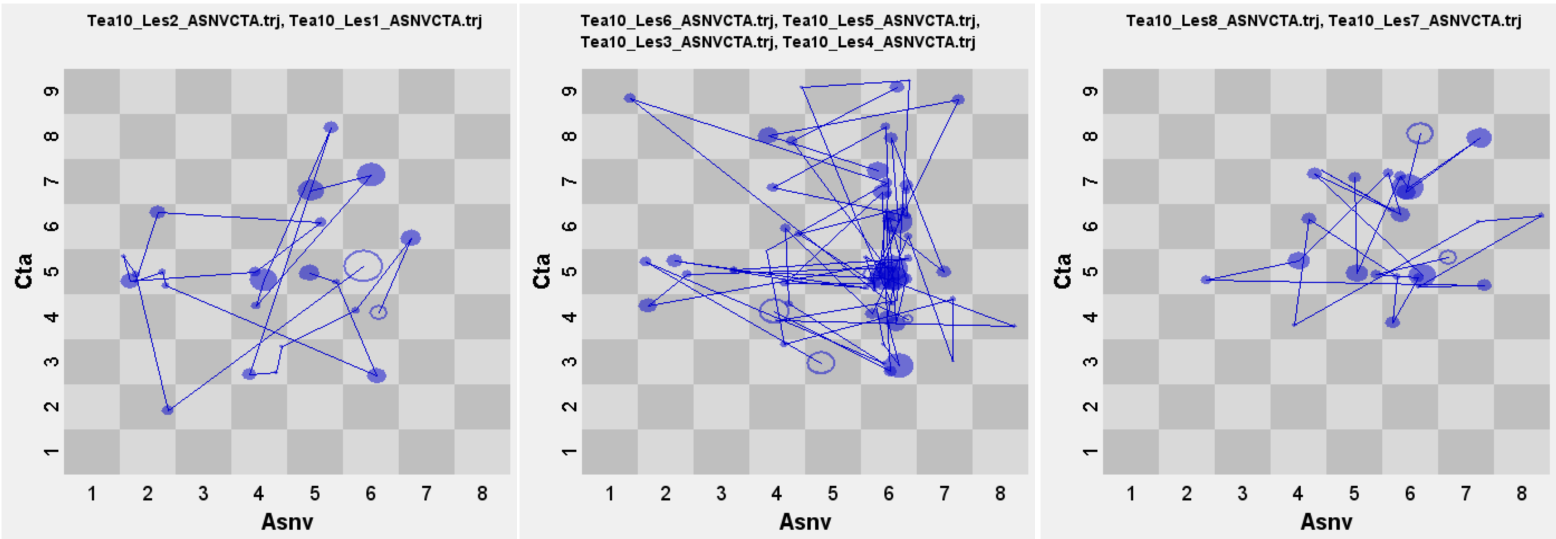
SSG of Teacher 9



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Figure D4

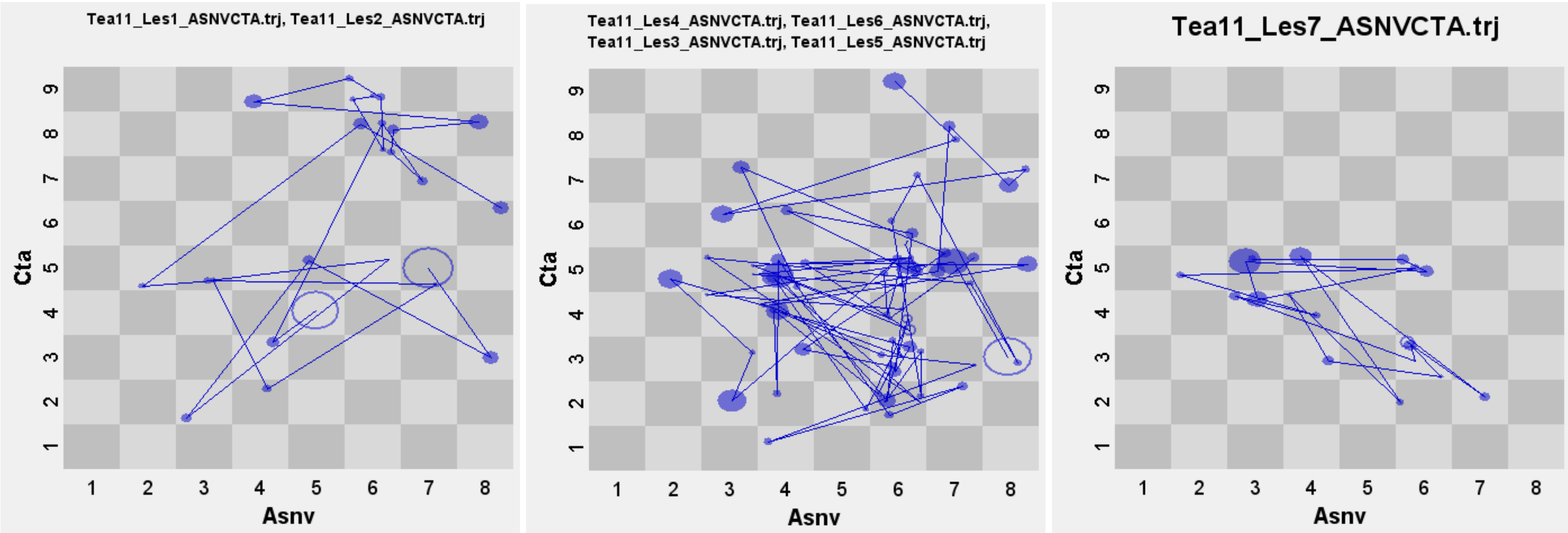
SSG of Teacher 10



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Figure D5

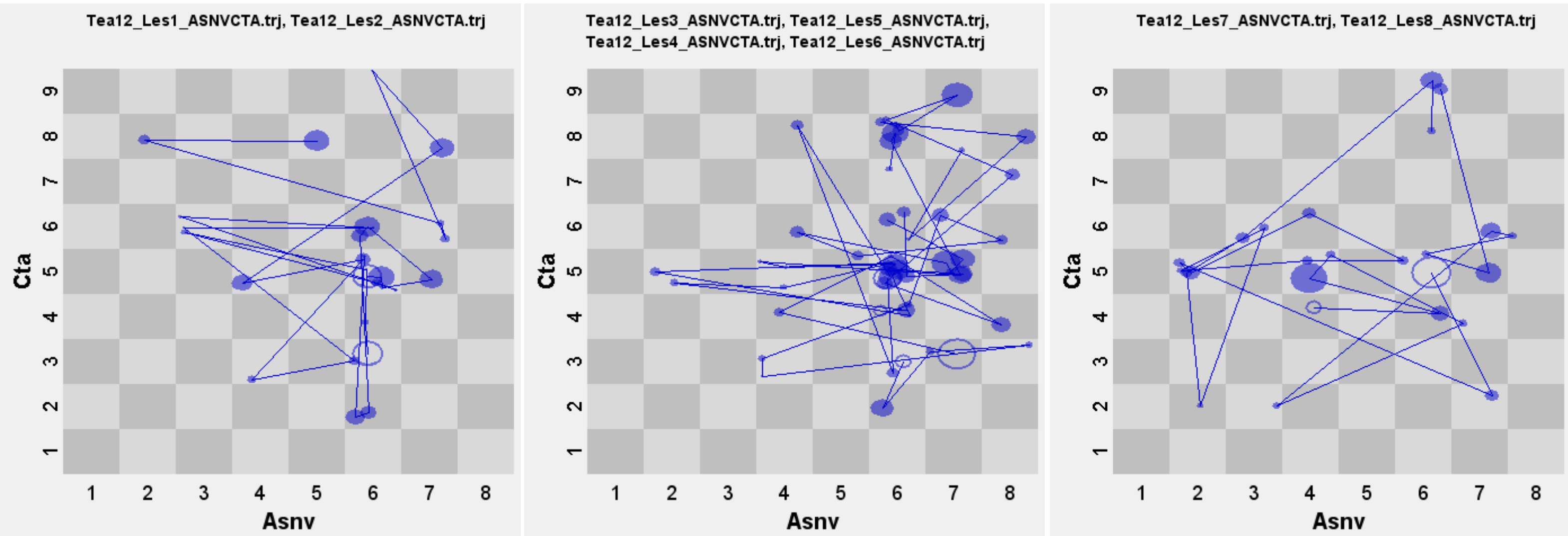
SSG of Teacher 11



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right

Figure D6

SSG of Teacher 12



Note. The research phases, pre-measurement, intervention, and post-measurement, are shown from left to right