

**The effect of music education on students' well-being. Empirical evidence from a field experiment**

Preliminary - work in progress

This draft: September 18, 2020

Vera Schramm  
University of Halle-Wittenberg

**Abstract** Our analyses, following the discussion above, address two central questions using a cultural capital framework. First, who participates in music both in and outside of school, and to what extent is such involvement stratified by social class, race/ethnic, and gender status? Second, and relative to the more central question discussed at the outset, do various forms of music involvement influence academic achievement, even after accounting for prior achievement, background statuses, and other educationally meaningful investments? Relatedly, to what extent might disparities in music involvement shape group-specific gaps in achievement that have been so well documented elsewhere?

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Music</b>	<b>5</b>
<b>3</b>	<b>Data</b>	<b>9</b>
3.1	The project . . . . .	9
3.2	selection process (and pre-treatment differences) . . . . .	10
3.3	Measuring LS in children . . . . .	10
3.4	Descriptive statistics . . . . .	12
<b>4</b>	<b>Econometrics</b>	<b>13</b>
4.1	Identification / Partial Poolin... . . . .	13
4.2	Estimation . . . . .	15
<b>5</b>	<b>Results</b>	<b>17</b>
5.1	Complete pooling . . . . .	17
5.2	No-pooling . . . . .	18
5.3	Partial pooling . . . . .	19
<b>6</b>	<b>Conclusion</b>	<b>22</b>
	<b>References</b>	<b>23</b>
<b>A</b>	<b>Figures</b>	<b>27</b>
<b>B</b>	<b>Tables</b>	<b>28</b>

## 1 Introduction

In German schools, and music are often considered less important than the typical hard subjects like math and science. Due to a lack of teachers, many classes are cancelled, of which 80% are in the subject of music. In Saxony we see ongoing efforts to eliminate the subject of music from the curriculum entirely. Furthermore, the quality of music lessons suffers from the fact that 80% of its teaching staff are foreign to the subject (Möller, 2017).

Music education experts are concerned about this development. According to them, music should not be regarded as a private matter. Regardless of their socioeconomic background, school children must have the opportunity to receive high level music education because it is as important for a proper education as literacy and mathematics (Gebert, 2018). Prof. Höppner, the General Secretary of the German Music Council (Generalsekretär des Deutschen Musikrats), said in an interview that music education helps to build stable self-esteem by learning to access ones own emotions (Stoverock, n.d.). He points out that the phase where music can shape a young person explicitly well is completed by the age of 13. That stresses the importance of high quality music education for students in pre- and secondary school.

The Federal Association of Music Education (Bundesverband Musikunterricht (BMU)) has set up the “Agenda 2030” to initiate an improvement in music education. Similar to Höppner (Stoverock, n.d.), they consider music education as valuable and essential for a social and cultural society. It encourages children to take responsibility and to increase their sense of self-determination (Bundesverband Musikunterricht (BMU), 2016, p. 2).

In views of this broad societal debate, empirical research has concentrated on the effect of music education on cognitive abilities (school grades and IQ in particular). Not so much effort was spent on observing the connection of music and well-being. This seems surprising because life satisfaction and happiness have become central research areas in the social sciences. My thesis addresses this research gap and investigates the effects of music education on children’s overall life satisfaction and on satisfaction in specific areas, namely satisfaction with the class, satisfaction with friends, and satisfaction with the situation at school. It analyzes music education in the classroom where fifth and sixth grade students have one additional hour of music education

per week. The project is called “klasse.im.puls” and it promotes the establishment of musical training in secondary schools in Bavaria. The program was implemented with the intention to give every child the opportunity to learn how to play an instrument. Additional positive outcomes were expected: an increase in self-confidence and social competence, as well as a reduction in violent behavior<sup>1</sup> My analysis focuses on the change in the overall life satisfaction and in satisfaction within specific areas reported by the students over the course of the project. The term of *life satisfaction* refers to a cognitive evaluation of a person’s reaction to his or her life in contrast to *affect*, an ongoing emotional reaction. Combined, LS and affect yield subjective well-being (Diener, 2009, p. 71). I will approach the problem by using a multi level model, that accounts for differences on the level of the individual student, on the level of the class and on the school level. I do this by using Bayesian inference. . .

The interest in life satisfaction as an outcome of the music project stems from the idea that higher values of LS come with many benefits. Among other positive correlates, adolescents reporting very high levels of LS are less likely to be affected by depression, anxiety, negative affect and social stress compared to adolescents with very low life satisfaction. Also, they achieve higher SEAs and demonstrate higher mean scores of school satisfaction (Gilman & Huebner, 2006, p. 316; Proctor, Linley, & Maltby, 2009, p. 928). These results go in line with a study by Suldo & Huebner (2004, p. 94) that shows that LS could be a moderating variable in predictions of the development of psycho-pathological behaviors. Low life satisfaction may be an indication for externalizing behavior problems in the future. When life satisfaction is on a higher level, those behavior problems are less likely to occur (Suldo & Huebner, 2004, p. 100) The authors conclude that life satisfaction might operate as a buffer against the development of subsequent externalizing behavior problems “in the face of stressful life events” (Suldo & Huebner, 2004, p. 101). Kim, Conger, Elder, & Lorenz (2003) add to the discussion that externalizing behavior problems in turn lead to more stressful life events. That reciprocal interrelation of stressful life events and externalizing problems (reported as delinquent behaviors) leads to unhealthy dynamics of increasing stress and behavior problems. If higher LS leads better coping mechanisms with stressful life events, these dependencies could be reduced. LS is also positively correlated

---

<sup>1</sup>For more information:

with children having higher measures of self-esteem, internal locus of control, and extraversion (Huebner, 1991a, p. 107). These features help in building a solid foundation for later life. On the other hand, dissatisfaction with life is associated with adolescents having poor mental health (like anxiety and neuroticism) and physical health and being exposed to a higher risk of considering or attempting suicide (Huebner, 1991a, p. 107; Valois, Zullig, Huebner, & Drane, 2004, p. 94). Furthermore, Zullig, Valois, Huebner, Oeltmann, & Drane (2001, pp. 284–185) show that adolescents reporting low levels of overall life satisfaction are more likely to use drugs and alcohol earlier in life and in higher amounts than adolescents with medium or high life satisfaction.

Considering the statement of Höppner, one would expect the project to positively effect students' life satisfaction. Evidence for this relation would lend credibility to the project and support its continuation. It could also stress the importance of music education and signal the Ministry of Culture to keep music education in the curriculum and work on its implementation in federal states alongside Bavaria. However, there will be no in-depth investigation of the analysis of any observations. Describing possible reasons for certain outcomes must be left to music and education experts. . .

The structure of my thesis is as follows: At first there will be an overview of the current literature on music effecting students' lives. Next, the project and the dataset will be explained. Descriptive statistics are presented alongside a detailed diagnostic on pre-treatment differences in the treatment group compared to the control group. The estimation strategy is presented in chapter four, following the results in the fifth chapter. Finally, chapter six concludes and discusses.

*Gute Einleitung bei Southgate & Roscigno (2009)*

## **2 Music**

With regard to the effect of music on students' lives, most of the researchers are interested in academic outcomes and intelligence among students who are actively involved in music. Generally, there is the predominant perception of a positive link between music and cognitive abilities. Osborne, McPherson, Faulkner, Davidson, & Barrett (2015, p. 14) observed improved math skills and higher subjective well-being scores in children that were part of a music project. He also found them to have a better self-control over impulsive behavior. Yang (2015) (p. 385), Wetter, Koerner, & Schwaninger (2008) (p. 372), Hille (2014) (p. 62), and Guhn, Emerson, & Gouzouasis (2019) (p. 316) present evidence that children playing music have better grades at school. But this conclusion is not very meaningful as all of these studies follow a correlational design and do not allow for causal inference. Nonetheless, the very optimistic and as we will see not quite realistic belief has prevailed that playing music causes children to achieve better results at school. This conclusion needs to be *revised*. In an extensive review of the available evidence concerning associations between music and cognitive abilities, the picture is not so clear any longer (Schellenberg & Weiss, 2013). Small associations between music training and mathematical ability in correlational and quasi-experimental studies might result from individual differences in general intellectual ability (p. 527). The available evidence simply indicates that high-functioning children (i.e., higher IQ, better performance in school) are more likely than other children to take music lessons and to perform well in mathematics and other tests of cognitive ability (p. 534). This fits with the Hille (2014) study where the outcome difference in cognitive skills between musically active and inactive children reduces greatly when holding constant observable characteristics. An other plausible interpretation of study outcomes that fail to detect a causal relationship comes from Wetter et al. (2008). He points out the possibility that the relationship is explained by affluent parents being more likely to afford music lessons for their child and thus, the socio-economic background may be the cause of higher performance at school. Despite the weakness of the above studies to draw causal inference, there is slight evidence that there may be a causal direction *from* music training *to* cognitive abilities in a study by Schellenberg (2004). He compares two treatment groups who receive piano lesson and voice lesson respectively to a control group in which the children have drama lessons. Random assignment to the different conditions allowed for inference that music lessons caused small

increases in cognitive abilities (namely larger increases in full-scale IQ). However, this does not preclude the possibility that high-functioning children are more likely driven to play music. The misconception of music being a predictor for academic achievement is also discussed by Southgate & Roscigno (2009) (p. 17) who states that music is rather a mediator, to some degree, of family background and student status. Results from a meta-analysis, suggest that music training does not reliably enhance children and young adolescents' cognitive or academic skills, and that previous positive findings were probably due to confounding variables, such as placebo effects and lack of random allocation of participants (Sala & Gobet, 2016, p. 64). The effect size was reduced in studies that applied a proper study design (random allocation of participants to the treatment group and comparison to an active control group). With respect to the mathematical outcomes, the only study comparing a music training to an active control group and with random allocation of the participants to the group (Mehr, Schachner, Katz, & Spelke, 2013) found a negative effect size. These considerations are in clear contrast to the popular perception that music training enhances any non-music related cognitive skill.

However positive outcomes due to music could be found outside academics. The Norwegian psychologist and musicologist Even Ruud who in 1978 was the head of the first music therapy training performed extensive studies on music and identity. He states that cultural activities, explicitly music, can “contribute to a feeling of quality of life and the subjective sense of health.” (p. 96). In an experimental study by Costa-Giomi (2004) a positive causal effect was detected of piano lessons on self-esteem [144] but she did not find an effect on math computation scores. One especially popular project that was conducted to improve children's lives is Venezuela's National Music Education Program “El Sistema”. It is a large scale social music education program established by José Abreu in the 1970s. 300,000 children are equipped with instruments every year and receive regular after-school lessons and are playing in orchestras. The initial goal was to prevent children from using drugs and being involved in violence and crime which was successfully achieved. Being in orchestras also enhanced social behavior of the students through greater concern for others and their own well-being (Uy, 2012, p. 13). However positive effects go far beyond keeping adolescents away from drugs and violence: El sistema teaches the participating students to “reflect and act upon the world in order to transform it” [7]. Playing in



an orchestra means joy, motivation, teamwork, the aspiration to success [6]. The students pick up management and organizational skills and responsibility due to many roles and rules that they need to follow to stay in the program [p. 10]. Also, being in an orchestra gives the students the chance to conceptualize themselves as part of something much larger and greater (p. 11) and they learn to express greater concern for others' and their well-being (p. 13). El Sistema became internationally popular and was replicated in several countries. Osborne et al. (2015) reviewed the outcome of El Sistema inspired projects in Australia and found significantly higher subjective well-being scores in the participating group (p. 14). Students from the music program also had better self-control over impulsive behavior (p. 15). However, other studies come to contrary conclusions and fail to show a significant effect of music participation on well-being, social skills, emotional intelligence or self-esteem (Portowitz, Lichtenstein, Egorova, & Brand, 2009, p. 121; Schellenberg, 2011, p. 190 association). Again, findings are diffuse and therefore interpretation of any results needs to be with caution.

The takeaway from this section is we must be very careful when drawing causal conclusion and we have to adjust the methods.

This paper complements existing literature by using different music indicators and estimation strategies to provide further insight on the relationship between music and education as well as the potential sources of endogeneity.

### 3 Data

## Pre-treatment differences

#### 3.1 The project

The data come from the years 2012-2014 when KIP was conducted in six different secondary schools (Mittel- und Realschulen) in Nuremberg. Generally, there are seven types of implementations of the KIP project: Band, wind players, strings, choir, percussion, strings and plucked instruments. Here, only band classes are observed. A school was able to apply for the project if certain criteria were fulfilled. For example, the school had to make sure that there was a full-time music teacher who was qualified to carry out the project. That person was trained to lead a band class and attended yearly meetings with other music teachers to share experiences. This way, it can be assured that the treatment in each school looked pretty much the same and there were no noteworthy differences in the way the teachers arranged the music lessons in the band classes. The three music lessons per week were split into a) Instrumental lessons in small groups, b) ensemble play, and c) general music education. Latter was similar to how music lesson instructed in the control group. In each school was a band class and a control class. The band classes received three music lessons per week, whereas the control classes only received two lessons of regular music education. Music lessons in the treatment classes differed significantly from that in the control group. In the band classes, the teachers had a more practical approach and there was less theoretical music education than in the control group. The students played concerts in regular intervals and were thus able to share what they had learned. In five different points of time, the students were asked to fill in a questionnaire:

wave 1: Beginning of 5th grade (pre-treatment) wave 2: Mid of 5th grade wave 3: end of 5th grade wave 4: mid of 6th grade wave 6: end of 6th grade

In total, 394 students were observed, however not all of them occur in each wave. Some questions were only asked in wave 1 (i.e. sex) which lead to the problem that there were no information about gender for students who did not attend in the first wave. Only 309 reported their gender in wave 1 which leads to 85 students for whom the gender is “unknown”. To not miss all of these

observations in regressions including gender, NAs were replaced by a category “unknown”. The same was done for the variable of migration background and language spoken at home. An other issue to point out is the comparably low number of observations in wave 5 in school 5.

Students had to rate their level of life satisfaction on a scale reaching from 0 (“completely unsatisfied”) to 10 (completely satisfied) for the question “How satisfied are you with your life overall?”

### **3.2 selection process (and pre-treatment differences)**

The decision who attended the treatment group was made before the entrance to grade 5. Parents could choose to put their child into a music class. Therefore, the experiment is not fully randomized. Children who were already into music in the first place might have been more likely to have chosen the music class. Also the decision can be a consequence of parental and socio-economic background of the student. To have a better idea of possible differences between the treatment and the control group, I computed standardized mean differences:

The standardized mean differences are not a reason of concern for the most part. For a major part of the observed variables, there are no notable differences between the treatment group and the control group. However, as already discussed, there are differences in prior musical knowledge. Among the treated, there are relatively more children who have music as a hobby. Of the 87 students who answered with “yes” to the question if music was their hobby, only 24 (SD=15.6) were assigned to the control group and the remaining 63 students attended the band classes (72%). Also, the duration of making music is very different in the two subgroups. In the Treatment group, 56/% are practicing at least 30 min, while in the control group that share is only 35/%. However, the model accounts for the individual musical background of the child. Therefore the fact that the control group and the treatment group does not a big issue.

### **3.3 Measuring LS in children**

To investigate one's quality of life, research uses both objective and subjective indicators. Typical objective indicators are the income levels, crime rates, and access to medical services – measures that are external and quantifiable. Subjective indicators on the other hand comprise subjective

evaluations of ones' individual life circumstances (Gilman & Huebner, 2000, p. 178). Only a modest relation between both measures was found which indicates that each approach carries unique information that are relevant for a comprehensive understanding of overall life quality (Veenhofen, 1996). Over time, measurement methods for subjective indicators have evolved leading to substantial growth in the life satisfaction research. However, for a long time, most of the measurements were designed to assess adults' life satisfaction. Only recently, investigating correlates of life satisfaction in adolescents has begun. One reason why life satisfaction research in children was put off for a long time are probably that measuring life satisfaction in children is more challenging than for adults. Instruments for assessing children's subjective life satisfaction reports have been less intensively developed which is probably due to the fact that a Likert-type ratings scale is more difficult to use for younger children than for adolescents (Chambers & Johnston, 2002, p. 28). For example, younger children or children with poorer readings skills are less able to respond appropriately to negative items on questionnaires and this effect biases the interpretation of children's responses (Marsh, 1986, p. 45). It is also common for children rating their subjective life satisfaction to show elevated extreme scores. As children become older, this tendency shriks and they are more capable of providing graded ratings in between the two extremes. These results have potentially substantial implications for the interpretation of self-report ratings form children *this tendency might have an erroneous and invalid impact on the interpretation of children's self-reports*(Chambers & Johnston, 2002, pp. 33–34). When dealing with self-reported life satisfaction in children, it is crucial that the respective child fully understands the question in order to give a valid response (Gluskie, 2012; Tomy, Fuller-Tyszkiewicz, Cummins, & Norrish, 2016). One must make sure that a child is old enough to know how to use a satisfaction scale. This requires children to distnace themselves from the current situation, cognitively evaluate their life satisfaction (considering all relevant areas of life) and rate the degree to which certain items on the scale apply to them. This requires abstract thinking, which children develop in early adolescent years (10-14 years) (Gluskie, 2012; Piaget, 1955, 1969). Gilman & Huebner (2000) have reviewed five different (both unidimensional and multidimensional) measurements explicitly developed to asses adolescents' life satisfaction.<sup>2</sup>

---

<sup>2</sup>The Students' Life Satisfaction Scale (Huebner, 1991b), the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), the Perceived Life Satisfaction Scale (Adelman, Taylor, & Nelson, 1989), the Comprehen-

The authors evaluated those measures in terms of validity and reliability and found all of the scales to be appropriate for research with adolescents [181-188]. The demographic characteristics of the available samples show that all of the adolescents observed were older than 12 years. Therefore it remains unclear if children younger than that age are able to report valid satisfaction. Another instrument was developed more recently by (Cummins & Lau, 2005) which is the Personal Well-being Index (PWI-SC). Again, studies demonstrated reliability for this instrument as well (Casas & Rees, 2015; Casas et al., 2011; Tomy & Cummins, 2011; Tomy, Stokes, Cummins, & Dias, 2019). But also in those studies, all of the adolescents were at least 12 years of age, mostly even older. There is only very little evidence on the psychometric properties of the PWI-SC for children below the age of 12. One of them is González-Carrasco, Casas, Malo, Viñas, & Dinisman (2016, p. 70) who applied the instrument for children as young as only 9 years and also found adequate fit of the data. On the other hand, Tomy et al. (2016) conducted a study with children aged 10-12 and concluded that subjective well-being data of children must be interpreted with caution. They also show that response bias towards the extreme positive end of a scale is higher with decreasing age. The authors do not recommend using the PWI-SC for children younger than 12 years. As for the specific sample, the PWI-SC did not serve as a valid instrument for measuring the SWB.

In conclusion, measuring life satisfaction in children is more challenging than for adults and is still in progress. It is advisable to check the validity and reliability of their data when testing children. Considering, the majority of the children from the KIP project are 10 year old in wave 1 (72% of those participating in wave 1), they might be just too young to give valid responses when asked about their life satisfaction. This must be kept in mind when evaluating the significance of the results.

### **3.4 Descriptive statistics**

---

sive Quality of Life Scale - School Version (Cummins, 1997; Gullone & Cummins, 1999), and the Multidimensional Student's Life Satisfaction Scale (Huebner, 1994)

## 4 Econometrics

An extensive amount of research was done on adult life satisfaction and on methods how to measure it. From that we know that not only is life satisfaction a result of life circumstances but it determines outcomes in several areas like health, ... (see Firsch 1999 for a review (suldo and huebner, B. 94))

Even though, the amount of research on

### 4.1 Identification / Partial Poolin...

“The challenge for identifying causal differences is the fact that...” - Ability - socio-economic characteristics - other background characteristics - we need to take into account selection effects - Bayesian because small sample

Gelman and Hill: Motivation for multilevel modelin: causal inference, prediction, descriptive modeling

Learning about treatment effects that vary Multilevel models allow us to study effects that vary by group (school!) is the intervention more effective in some schools than others? why? Only few observtions per school, multilevel modeling allows us to estimat these interactions to the extent supported by the data

Using all the data to perform inference for groups with small sample size classical estimation just using the local information can be essentially useless if the samle size is small in the group - a classical regression ignoring group indicators can be misleading in ignoring group-level variation (no pooling, complete pooling)

Prediction predicting outcomes of new cases. but what if the data vary by group? Then we can make predictions for new units in existing groups or in new groups

inference should include the factors used in the design of data collection. without being overwhelmed with the problems of overfittin that arise form applying least squares or maximum likelihood

Traditional alternatives to multilevel modeling are *complete pooling* in which differences between groups are ignored and *no pooling* in which data from different sources are analyzed separately. Elreath: “If we ignore the clusters, assigning the same intercept to each of them, then we risk ignoring important variation in baseline survival. This variation could mask association with other variables. If we instead estimate a unique intercept for each cluster, using a dummy variable for each tank, we instead practice anterograde amnesia.” p416 Both these approaches have problems: no pooling ignores information and can give unacceptably variable inference, and complete pooling suppresses variation that can be important or even the main goal of a study. The extreme alternatives can in fact be useful as preliminary estimates, but ultimately we prefer the partial pooling that comes out of a multilevel analysis.

Multilevel models are extensions of regression in which data are structured in groups and coefficients can vary by group

varying intercept and varying slope indicators for groups: varying-intercept model p237 varying slopes are interactions between the continuous predictor  $x$  and the group indicators First step: set up a regression with varying coefficients second step: set up a regression model for the coefficients themselves

McElreath Multilevel models remember features of each cluster in the data as they learn about all of the clusters. Depending among clusters, which is learned from the data as well, the model pools information across clusters. This pooling tends to improve estimates about each cluster. This improved estimation leads to several, more pragmatic sounding, benefits of the multilevel approach: 1. Improved estimates for repeat sampling 2. Improved estimates for imbalance in sampling 3. Estimates of variation 4. Avoid averaging, retain variation

p416 after all, tanks are different but each tank does help us estimate survival in the other tanks. So it doesn't make sense to forget entirely, moving from one tank to another. A multilevel model, in which we simultaneously estimate both an intercept for each tank and the variation among tanks, is what we want. ... This adaptive learning is the absence of amnesia discussed at the start of the chapter. When what we learn about each cluster informs all the other clusters, we learn the prior simultaneously to learning the intercepts. regularizing priors

p422 Complete pooling: This means we assume that the population of ponds is invariant, the same as estimating a common intercept for all ponds No pooling: This means we assume that each pond tells us nothing about any other pond. This is the model with amnesia. Partial pooling: This means using an adaptive regularizing prior, as in the previous section (more details on subsequent pages)

p50 (34) In future chapters, there will be more parameters in your models. In statistical modeling, many of the most common questions we ask about data are answered directly by parameters:

- \* What is the average difference between treatment groups?
- \* How strong is the association between a treatment and an outcome?
- \* Does the effect of the treatment depend upon a covariate?
- \* How much variation is there among groups?

You'll see how these questions become extra parameters inside the distribution function we assign to the data. For every parameter you intend your Bayesian machine to consider, you must provide a distribution of prior plausibility, its *PRIOR*.

## 4.2 Estimation

I simultaneously investigate the treatment in six different schools. The analysis is restricted to wave 1 (pre treatment) and wave 2 (post treatment). The regression equation is:

$$y_{ijt} = \beta_0 + \beta_{1j[i]}T_i + \sum_{l=2}^5 \delta_{lj[i]}(T_i \times \text{Period}_l) + \lambda_t + \mu_{j[i]} + \alpha_i + \epsilon_{ijt},$$

where

- $y_{ijt}$  is the outcome of pupil  $i$  in school  $j$  at time  $t$ .
- $T_i$  is a binary indicator of whether or not pupil  $i$  is in the treatment group. We model a fixed effect for the treatment group because the treatment takes place at the aggregate level (class). Therefore, bias in the estimate of the treatment effect may result from unobserved factors at the class level.
- $\delta_{lj[i]}$  is the treatment effect in wave  $l$ . The effect is allowed to vary across schools.
- $\lambda_t$  denotes period fixed effects.



- $\alpha_i$  models individual-specific unobserved factors. Individual heterogeneity is model because we have repeated observations for the same pupils over time.
- $\varepsilon_{it}$  is an error term.

(Simulation?)

For doing *blah* I use *blub* [ref]

For each outcome variable... First I do no pooling, then I do pooling Then I do Bayesian

## 5 Results

As described in Section 4, estimation biases resulting from selection into treatment take place at two stages: The initial decision to take up music lessons and the decision not to give up until age 17... At such a young age, the choice of a long-term extracurricular activity such as music is strongly determined by the parents. For the parents, however, we observe a large number of characteristics, in particular their socio-economic status, personality, involvement with the child's education, and taste for the arts... (Hille, 2014, p. 65) NOTE: I have to observe parents ("parental background differences") NOTE: Results in Hille (2014) might be driven by unobserved heterogeneity... unobserved individual characteristics could still determine the decision to keep on playing music until age 17 rather than giving up earlier (p. 67). NOTE: Sensitivity of Hille (2014) results to reverse causality by performing mediation analysis in which we estimate the correlation between music practice and outcome  $p$ , while subsequently controlling for all outcomes  $q$  rather than  $p$  (p. 67). How do I address the issue of reverse causality? Three challenges should determine the agenda of future research on this agenda. 1. separate the influence of parental and individual background from that of music (identify a variable that increases the likelihood to learn a musical instrument without affecting the development of skills.) 2. answer the question of the extent to which extracurricular activities are substitutable (substitutes vs. compliments). 3. long-term effects of music training on outcomes such as labor market success or life satisfaction

### 5.1 Complete pooling

Table 1: (#tab:complete pooling)Estimates complete pooling

wave	complete	lower	upper
1	-0.0171144	-0.3985530	0.3643241
2	-0.1543548	-0.5153323	0.2066226
3	0.3483488	-0.0247137	0.7214114
4	-0.3285619	-0.7027461	0.0456223

## 5.2 No-pooling

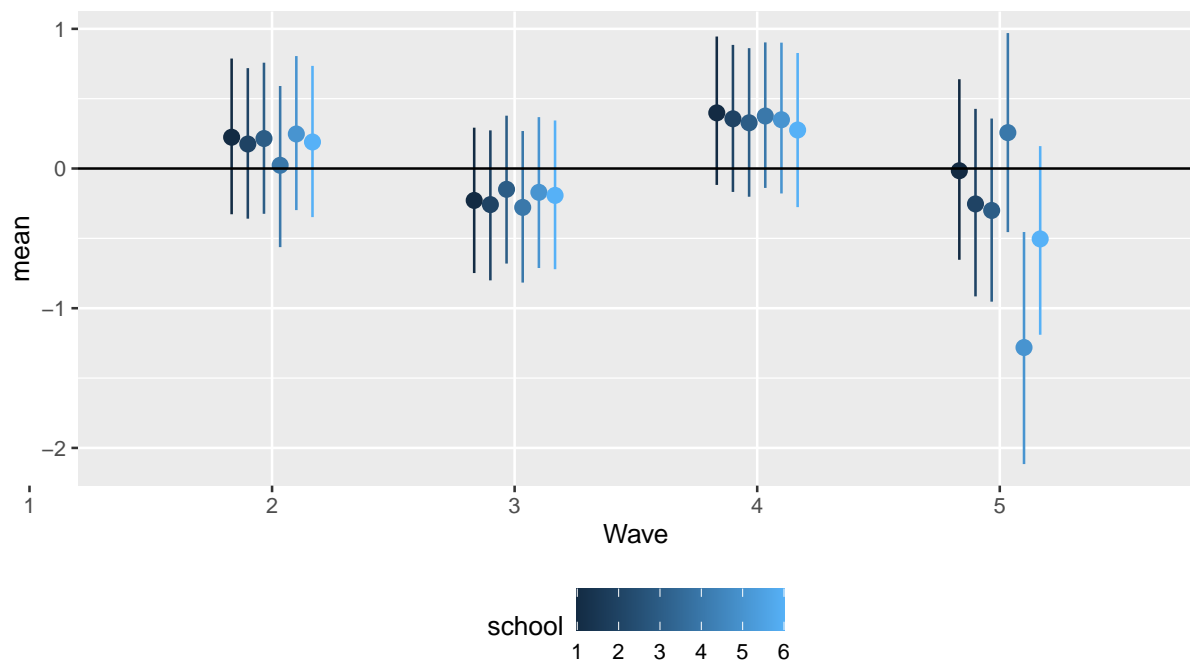
Table 2: Estimates no-pooling

school	wave	estimate	lower	upper
1	1	-0.2126153	-0.5037749	0.0785443
1	2	-0.2338833	-0.5205275	0.0527609
1	3	-0.1318997	-0.4227986	0.1589992
1	4	-0.1845208	-0.4722545	0.1032128
2	1	-0.1625532	-0.6963053	0.3711989
2	2	-0.1838212	-0.7130579	0.3454154
2	3	-0.0818376	-0.6153290	0.4516537
2	4	-0.1344587	-0.6647848	0.3958674
3	1	0.4478898	-0.0819770	0.9777565
3	2	0.4266218	-0.0987296	0.9519731
3	3	0.5286054	-0.0010007	1.0582114
3	4	0.4759843	-0.0504565	1.0024250
4	1	-0.7582930	-1.2949003	-0.2216857
4	2	-0.7795610	-1.3116529	-0.2474691
4	3	-0.6775774	-1.2139240	-0.1412308
4	4	-0.7301985	-1.2633798	-0.1970171
5	1	-0.5744479	-1.1097624	-0.0391334
5	2	-0.5957159	-1.1265150	-0.0649168
5	3	-0.4937323	-1.0287861	0.0413215
5	4	-0.5463534	-1.0782419	-0.0144648
6	1	-0.7582930	-1.2907278	-0.2258581
6	2	-0.7795610	-1.3074804	-0.2516415
6	3	-0.6775774	-1.2097515	-0.1454033
6	4	-0.7301985	-1.2592074	-0.2011896

### 5.3 Partial pooling

Table 3: (#tab:partial pooling)Estimates partial pooling

school	wave	estimate	lower	upper	Q10	Q90	mean
7	1	NA	NA	NA	-	0.6909753	0.1806651
					0.3164921		
7	2	NA	NA	NA	-	0.2894317	-
					0.6898667		0.2062209
7	3	NA	NA	NA	-	0.8530004	0.3448464
					0.1488093		
7	4	NA	NA	NA	-	0.2854370	-
					0.9917535		0.3541457



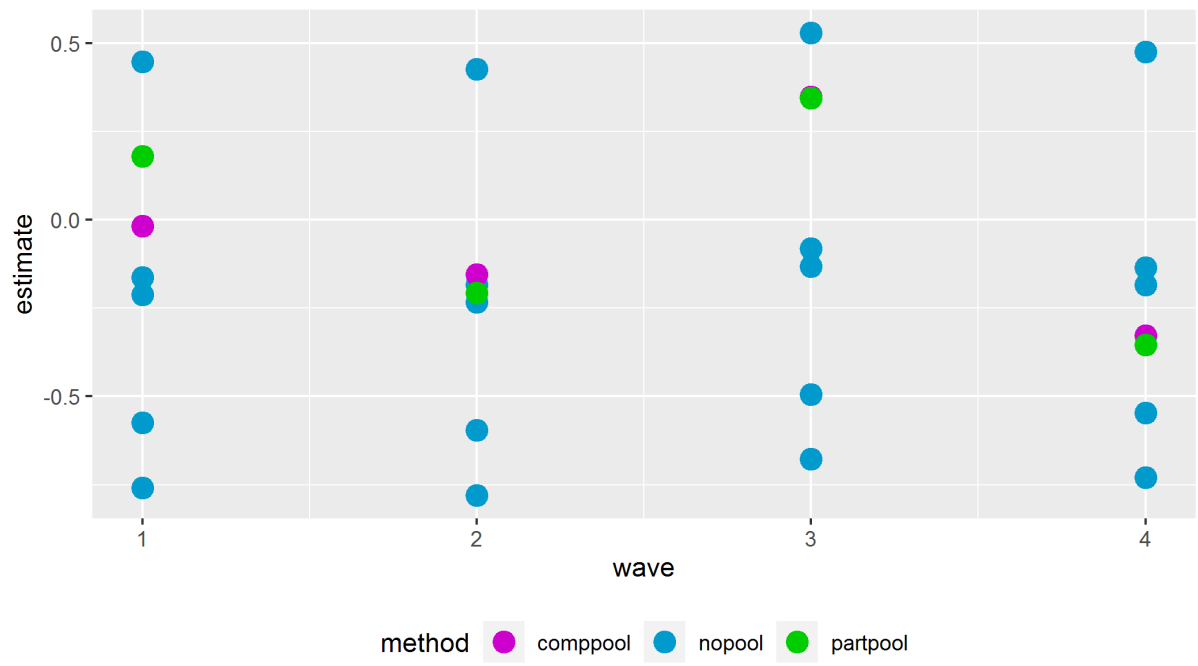


Table 4: (#tab:compare methods)Estimates generated by different methods

school	wave	estimate	lower	upper	Q10	Q90	method
1	1	-	-	0.0785443	NA	NA	nopool
		0.2126153	0.5037749				
1	2	-	-	0.0527609	NA	NA	nopool
		0.2338833	0.5205275				
1	3	-	-	0.1589992	NA	NA	nopool
		0.1318997	0.4227986				
1	4	-	-	0.1032128	NA	NA	nopool
		0.1845208	0.4722545				
2	1	-	-	0.3711989	NA	NA	nopool
		0.1625532	0.6963053				
2	2	-	-	0.3454154	NA	NA	nopool
		0.1838212	0.7130579				
2	3	-	-	0.4516537	NA	NA	nopool
		0.0818376	0.6153290				
2	4	-	-	0.3958674	NA	NA	nopool
		0.1344587	0.6647848				
3	1	0.4478898	-	0.9777565	NA	NA	nopool
		0.0819770					
3	2	0.4266218	-	0.9519731	NA	NA	nopool
		0.0987296					
3	3	0.5286054	-	1.0582114	NA	NA	nopool
		0.0010007					
3	4	0.4759843	-	1.0024250	NA	NA	nopool
		0.0504565					
4	1	-	-	-	NA	NA	nopool
		0.7582930	1.2949003	0.2216857			
4	2	-	-	-	NA	NA	nopool
		0.7795610	1.3116529	0.2474691			
4	3	-	-	-	NA	NA	nopool
		0.6775774	1.2139240	0.1412308			
4	4	-	-	-	NA	NA	nopool
		0.7301985	1.2633792	10.1970171			
5	1	-	-	-	NA	NA	nopool
		0.5744479	1.1097624	0.0391334			

## 6 Conclusion

## References

- Adelman, H. S., Taylor, L., & Nelson, P. (1989). Minors dissatisfaction with their life circumstances. *Child Psychiatry & Human Development*, 20(2), 135–147. <https://doi.org/10.1007/bf00711660>
- Bundesverband Musikunterricht (BMU). (2016). *Grundsatzpapier*. Retrieved from <https://www.bmu-musik.de/ueber-uns/positionen/agenda-2030-bmu-positionen-9-2016.html>
- Casas, F., & Rees, G. (2015). Measures of children's subjective well-being: Analysis of the potential for cross-national comparisons. *Child Indicators Research*, 8(1), 49–69. <https://doi.org/10.1007/s12187-014-9293-z>
- Casas, F., Sarriera, J. C., Alfaro, J., González, M., Malo, S., Bertran, I., ... Valdenegro, B. (2011). Testing the personal wellbeing index on 1216 year-old adolescents in 3 different countries with 2 new items. *Social Indicators Research*, 105(3), 461–482. <https://doi.org/10.1007/s11205-011-9781-1>
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in childrens use of rating scales. *Journal of Pediatric Psychology*, 27(1), 27–36. <https://doi.org/10.1093/jpepsy/27.1.27>
- Costa-Giomi, E. (2004). Effects of three years of piano instruction on children's academic achievement, school performance and self-esteem. *Psychology of Music*, 32(2), 139–152. <https://doi.org/10.1177/0305735604041491>
- Cummins, R. A. (1997). *Manual for the comprehensive quality of life scale – student (grades 7-12): ComQol-s5* (5th ed.). Malbourne: Deakin University, School of Psychology.
- Cummins, R. A., & Lau, A. L. D. (2005). *Personal wellbeing index - school children (PWI-SC)* (3rd ed.). Australian Centre on Quality of Life, School of Psychology, Deakin University, Australia.
- Diener, E. (2009). *Culture and well-being* (T. Moum, M. A. G. Sprangers, J. Vogel, R. V. C. Michalos, E. Diener, & W. G. and, Eds.). <https://doi.org/10.1007/978-90-481-2352-0>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71–75. [https://doi.org/10.1207/s15327752jpa4901\\_13](https://doi.org/10.1207/s15327752jpa4901_13)
- Gebert, S. (2018). *Musische erziehung ist keine privatangelegenheit*. Retrieved from [https://www.deutschlandfunk.de/musikunterricht-an-schulen-musische-erziehung-ist-keine.680.de.html?dram:article\\_id=419333](https://www.deutschlandfunk.de/musikunterricht-an-schulen-musische-erziehung-ist-keine.680.de.html?dram:article_id=419333)
- Gilman, R., & Huebner, E. S. (2000). Review of life satisfaction measures for adolescents. *Behaviour Change*, 17(3), 178–195. <https://doi.org/10.1375/bech.17.3.178>
- Gilman, R., & Huebner, E. S. (2006). Characteristics of adolescents who report very high life satisfaction. *Journal of Youth and Adolescence*, 35(3), 293–301. <https://doi.org/10.1007/s10964-006-9036-7>
- Gluskie, A. L. (2012). *Subjective wellbeing in children* (PhD thesis). Deakin University.



- González-Carrasco, M., Casas, F., Malo, S., Viñas, F., & Dinisman, T. (2016). Changes with age in subjective well-being through the adolescent years: Differences by gender. *Journal of Happiness Studies*, 18(1), 63–88. <https://doi.org/10.1007/s10902-016-9717-1>
- Guhn, M., Emerson, S. D., & Gouzouasis, P. (2019). A population-level analyses of associations between school music participation and academic achievement. *Journal of Educational Psychology*, 112(2), 308–328. <https://doi.org/http://dx.doi.org/10.1037/edu0000376>
- Gullone, E., & Cummins, R. (1999). The comprehensive quality of life scale: A psychometric evaluation with an adolescent sample. *Behaviour Change*, 16, 127–139.
- Hille, J., Adrian; Schupp. (2014). How learning a musical instrument affects the development of skills. *Economics of Education Review*, 44, 56–82. <https://doi.org/http://dx.doi.org/10.1016/j.econedurev.2014.10.007>
- Huebner, E. S. (1991a). Correlates of life satisfaction in children. *School Psychology Quarterly*, 6(2), 103–111. <https://doi.org/https://doi.org/10.1037/h0088805>
- Huebner, E. S. (1991b). Initial development of the students life satisfaction scale. *School Psychology International*, 12(3), 231–240. <https://doi.org/10.1177/0143034391123010>
- Huebner, E. S. (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. *Psychological Assessment*, 6(2), 149–158. <https://doi.org/10.1037/1040-3590.6.2.149>
- Kim, K. J., Conger, R. D., Elder, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child Development*, 74(1), 127–143. <https://doi.org/10.1111/1467-8624.00525>
- Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22(1), 37–49. <https://doi.org/10.1037/0012-1649.22.1.37>
- Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. *PLoS ONE*, 8(12), e82007. <https://doi.org/10.1371/journal.pone.0082007>
- Möller, T. (2017). *Ausverkauf musikalischer Bildung?* Retrieved from [https://www.deutschlandfunk.de/musikunterricht-in-der-schule-ausverkauf-musikalischer.1992.de.html?dram:article\\_id=382783](https://www.deutschlandfunk.de/musikunterricht-in-der-schule-ausverkauf-musikalischer.1992.de.html?dram:article_id=382783)
- Osborne, M. S., McPherson, G. E., Faulkner, R., Davidson, J. W., & Barrett, M. S. (2015). Exploring the academic and psychosocial impact of el sistema-inspired music programs within two low socio-economic schools. *Music Education Research*, 18(2), 156–175. <https://doi.org/10.1080/14613808.2015.1056130>
- Piaget, J. (1955). *The construction of reality in the child*. London: Routledge; Keagan Paul.
- Piaget, J. (1969). *The child's concept of time*. London: Routledge; Keagan Paul.
- Portowitz, A., Lichtenstein, O., Egorova, L., & Brand, E. (2009). Underlying mechanisms linking music education and cognitive modifiability. *Research Studies in Music Education*, 31(2), 107–128. <https://doi.org/10.1177/1321103x09344378>

- Proctor, C., Linley, P. A., & Maltby, J. (2009). Very happy youths: Benefits of very high life satisfaction among adolescents. *Social Indicators Research*, 98(3), 519–532. <https://doi.org/10.1007/s11205-009-9562-2>
- Sala, G., & Gobet, F. (2016). When the musics over. Does music skill transfer to childrens and young adolescents cognitive and academic skills? A meta-analysis. *Educational Research Review*, 20, 55–67. <https://doi.org/10.1016/j.edurev.2016.11.005>
- Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, 15(8), 511–514. <https://doi.org/https://doi.org/10.1111/j.0956-7976.2004.00711.x>
- Schellenberg, E. G. (2011). Music lessons, emotional intelligence, and IQ. *Music Perception: An Interdisciplinary Journal*, 29(2), 185–194. <https://doi.org/10.1525/mp.2011.29.2.185>
- Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In *The psychology of music* (pp. 499–550). <https://doi.org/10.1016/b978-0-12-381460-9.00012-2>
- Southgate, D. E., & Roscigno, V. J. (2009). The impact of music on childhood and adolescent achievement. *Social Science Quarterly*, 90(1), 4–21. <https://doi.org/10.1111/j.1540-6237.2009.00598.x>
- Stoverock, K. (n.d.). *Ein jahrzehntelanges versagen der bildungspolitik*. Retrieved from <https://themen.miz.org/fokus-musikunterricht/interview-hoeppner>
- Suldo, S. M., & Huebner, E. S. (2004). Does life satisfaction moderate the effects of stressful life events on psychopathological behavior during adolescence? *School Psychology Quarterly*, 19(2), 93–105. <https://doi.org/10.1521/scpq.19.2.93.33313>
- Tomyn, A. J., & Cummins, R. A. (2011). The subjective wellbeing of high-school students: Validating the personal wellbeing indexSchool children. *Social Indicators Research*, 101(3), 405–418. <https://doi.org/10.1007/s11205-010-9668-6>
- Tomyn, A. J., Fuller-Tyszkiewicz, M. D., Cummins, R. A., & Norrish, J. M. (2016). The validity of subjective wellbeing measurement for children: Evidence using the personal wellbeing indexSchool children. *Journal of Happiness Studies*, 18(6), 1859–1875. <https://doi.org/10.1007/s10902-016-9804-3>
- Tomyn, A. J., Stokes, M. A., Cummins, R. A., & Dias, P. C. (2019). A rasch analysis of the personal well-being index in school children. *Evaluation & the Health Professions*, 43(2), 110–119. <https://doi.org/10.1177/0163278718819219>
- Uy, M. (2012). Venezuela's national music education program el sistema: Its interactions with society and its participants' engagement in praxis. *Music & Arts in Action*, 4(1), 5–21. Retrieved from <http://musicandartsinaction.net/index.php/maia/article/view/elsistema>
- Valois, R. F., Zullig, K. J., Huebner, E. S., & Drane, J. W. (2004). Life satisfaction and suicide among high school adolescents. *Social Indicators Research*, 66(1/2), 81–105. <https://doi.org/10.1023/b:soci.0000007499.19430.2f>
- Wetter, O. E., Koerner, F., & Schwaninger, A. (2008). Does musical training improve school performance? *Instructional Science*, 37(4), 365–374. <https://doi.org/10.1007/s11251-008-9052-y>

- Yang, P. (2015). The impact of music on educational attainment. *Journal of Cultural Economics*, 39(4), 369–396. <https://doi.org/10.1007/s10824-015-9240-y>
- Zullig, K. J., Valois, R. F., Huebner, E. S., Oeltmann, J. E., & Drane, J. W. (2001). Relationship between perceived life satisfaction and adolescents' substance abuse. *Journal of Adolescent Health*, 29(4), 279–288. [https://doi.org/10.1016/s1054-139x\(01\)00269-5](https://doi.org/10.1016/s1054-139x(01)00269-5)

Example of nice appendix in Hille (2014)

## **A Figures**

## B Tables

Table 5: Standardized mean differences

	0	1	SMD
n	154	163	
lsat (mean (SD))	7.50 (2.39)	7.11 (2.70)	0.155
female (%)			0.159
0	82 (53.2)	88 (54.0)	
1	70 (45.5)	69 (42.3)	
unknown	2 ( 1.3)	6 ( 3.7)	
sat_friends (mean (SD))	8.70 (1.82)	8.99 (1.82)	0.157
sat_class (mean (SD))	7.74 (2.06)	7.42 (2.00)	0.156
note.deu (mean (SD))	2.53 (0.62)	2.59 (0.61)	0.097
note.mat (mean (SD))	2.46 (0.74)	2.56 (0.73)	0.128
note.mus (mean (SD))	1.91 (0.92)	1.75 (0.77)	0.192
hob.mus.making = Ja (%)	24 (15.6)	63 (38.7)	0.537
dauer.mus.lis (mean (SD))	2.61 (1.02)	2.51 (1.04)	0.097
dauer.mus.making (%)			0.450
gar nicht	94 (64.8)	70 (44.0)	
ungefŠhr 30 Min. 1-2 Std.	34 (23.4)	54 (34.0)	
ungefŠhr 3-4 Std.	13 ( 9.0)	29 (18.2)	
ungefŠhr mehr als 4 Std.	3 ( 2.1)	3 ( 1.9)	
play.instr = Ja (%)	1 ( 0.7)	3 ( 1.9)	
mus.active (mean (SD))	41 (30.1)	53 (34.6)	0.096
migback (%)	0.33 (0.47)	0.41 (0.49)	0.165
Native			0.151
Immigrant	102 (67.1)	98 (60.1)	
Unknown	49 (32.2)	63 (38.7)	
	1 ( 0.7)	2 ( 1.2)	