

Do grades reflect the development of excellence in music students? The prognostic validity of entrance exams at universities of music

Musicae Scientiae
2014, Vol. 18(2) 232–248

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1029864914530394

msx.sagepub.com



Anna Wolf and Reinhard Kopiez

Hanover University of Music, Germany

Abstract

For successful admission to universities of music, prospective students have to pass entrance exams which assess the quality of their music-making and, in some instances, their abilities in music theory and aural training. However, only very little is known about the validity of such grades and the skill development of expectant professional musicians. For the first time, the present study analyzes the long-term development of grades of students in a bachelor degree program ($N = 63$) over a period of 3 years. As the grades were neither interval-scaled nor normally distributed, a logistic regression analysis was calculated to quantify the predictability of final grades in the main instrument and music theory based on the respective entrance exam grades. The prognostic validity was low for grades in both the main instrument and music theory/aural training (Nagelkerke $R^2 = .08$ and $.01$). This result can be mainly attributed to two explanations: First, university grades are often inflated and subject to ceiling effects. Second, the curricula in music theory contain such variety that it is impossible to define what all students across the board learn. In order to better define expertise skill development in students, we suggest initiating regular competency-based assessments for professional musicians at a pre-collegiate and collegiate level.

Keywords

ear training, entrance exams, music theory, prognostic validity, assessment

Introduction

In the process of becoming a musical expert, one has to master an ample variety of music related skills (McPherson, Bailey & Sinclair 1997; McPherson & Gabrielsson, 2002). Musical ability is based on aural, cognitive, technical, and learning skills as well as those skills acquired in musicianship and performance (Hallam, 1998). Aural and cognitive skills are mainly defined as resulting from expertise in music theory and ear training.

Corresponding author:

Anna Wolf, Hanover University of Music, Drama and Media, Emmichplatz 1, 30175 Hanover, Germany.

Email: anna.wolf@hmtm-hannover.de

The importance of music theory and ear training in music education

Expertise in music theory and ear training are defined by the relevant chapters of *The Oxford Companion to Music* (Fallows, 2011; Fry & Spencer, 2011). The aim of ear training is “to improve communication between the ear and the brain, thus improving the listener’s conscious and intellectual grasp of what the ear hears” (Fry & Spencer, 2011, para. 1). The main purpose of classes in ear training is to enable students to transform the auditory percept into a (rudimentary) representation of the score. The encyclopedia articles mentioned above stress the importance of a high degree of expertise in the relevant skill. “No performer, teacher, or leader of an ensemble could function properly without a high degree of aural perception” (Fry & Spencer, 2011, para. 3) and “[music theory] promises important understanding for any thinking musician” (Fallows, 2011, para. 4). These statements rest upon the fact that anyone who works with music and musical scores (either written or imagined) needs to match both representations of information. This is only possible if the individual understands both the musical and the notated representation (Hallam, 1998).

It needs to be kept in mind that such statements originate from researchers who dedicate their work to music theory and the teaching of it. The proclaimed effect in these statements is not always a prerequisite to music-making, since many musically active people have never explicitly learned music theory (i.e., pop musicians and many amateur musicians). Therefore, measuring the learning progress in music theory and its impact on musical performance will improve our understanding of these factors in musical engagements. The influence of music theory and ear training on musical skills has been explicitly and implicitly assessed in the studies on musical performance and music education below.

McPherson et al. (1997) validated a five-factor model relevant for an instrumentalist’s musical performance. These factors are sight-reading, performing rehearsed music, playing from memory (after memorizing the written score), playing by ear, and improvising. The associations between all five factors showed significant correlation coefficients between $r = .64$ and $r = .77$. Additional data on four post-hoc defined factors were gathered: early exposure, enriching activities, length of study, and quality of study. Length of study included one item about music theory (‘learning AMEB theory’) and enriching activities contained items on the frequency of composing and the election of classroom music. These were the only items that have to do with theory-related knowledge used in music performance.

As the five-factor model itself constitutes a balance of aural, visual and creative skills of musical performance, skills such as ear training and music theory are integrated in the five performance factors as well. Especially playing by ear and improvising, which were correlated by $r = .77$, cannot be acquired without connecting musical imagery and the instrument-dependent motor skills: In both tasks, the musician had to play a given or an invented musical representation, which would have not been possible if the sound had not been “understood” in terms of auditory imagery. For sight-reading, playing by memory, and performing rehearsed music, theoretical knowledge might seem less important. Nonetheless, musically sensible chunking, the recognition of repetitions, and the development of imagined sound representations of the piece are very beneficial competencies for playing music. While the exact influence of competencies in music theory and ear training on playing music still has to be identified, most (semi-)professional musicians are musically literate to some degree due to their long-standing training. Musicianship cannot be achieved by mechanically transforming notes into sound—musical literacy is characterized by a person’s ability to use notation for “thinking in sound” (McPherson & Gabrielsson, 2002).

Brodsky, Henik, Rubinstein, and Zorman (2003) examined notational audiation in which musicians produced musical imagery from a musical notation. This is an expendable but

beneficial skill for music-making that is assumed to be possessed by many musicians. Participants were supposed to study a musical score and then distinguish between target (same) and lure (different) tunes. However, results suggest that about half of the participating musicians were not as well-versed as necessary to perform above the level of chance in this signal detection task. It can therefore be concluded that notational audiation requires many hours of training and that many musicians lack this aspect of musical imagery despite being expert musicians.

An in-depth study on the skills predicting the sight-reading of previously unknown music has revealed the complex conjunctions of the skills best predicting the quality of playing a given piece (Kopiez & Lee, 2006, 2008). General cognitive skills, such as working memory capacity and short-term memory, seem to play a less important role. Elementary cognitive skills have a greater impact on sight-reading music: For instance, a person's speed of information processing influences their achievements in sight-reading (Kopiez & Lee, 2006, p. 117). Amongst the expertise-related skills, sight-reading expertise acquired by the age of 15 is the most relevant factor and presents a critical time window for successful sight-reading in later years. Another important predictor is the speed of trilling with two fingers, which is a skill both dependent on and independent of practice due to its high correlations with both solo expertise and tapping speed.

However, the precise impact of the independent variables on the quality of the sight-reading task depends on the difficulty of the played piece. For easier pieces, inner hearing and general pianistic skills are sufficient training for sight-reading, and they could be practiced by "playing by ear" and "improvising" from McPherson's five-factor model (McPherson et al., 1997). For the more difficult pieces, trilling speed, sight-reading expertise by the age of 15 and inner hearing are good predictors (Kopiez & Lee, 2006). A model independent of the difficulty of the played piece still can explain 60% of the variance of sight-reading when considering the following four factors: trilling speed, sight-reading expertise by the age of 15, speed of information processing, and inner hearing. These factors again stress the interaction of practice-related and practice-unrelated components. Evidently, in terms of music education this study has ramifications for sight-reading training during early adolescence and the training of pattern detection and chunking of notational elements (Kopiez & Lee, 2008).

Assessment of skills in music theory and ear training

While a precise knowledge and application of aural skills are necessary components of musicianship, there have not been objective methods to determine a person's state of knowledge in music theory and aural skill. Recently, there has been some research on the education of musical skills in children and adolescents (i.e., KoMus—Competence model for music, 2009; National Assessment of Educational Progress, 2008). Several countries have tried to develop professional study and test procedures for music theory and ear training, such as the USA with its Advanced Placement Music Theory exam (College Board, 2012) and the UK with tests by the Associated Board of the Royal Schools of Music (2012). For the latter, abilities in music theory and ear training can be trained and tested within a well-developed curriculum, and ear training tests are included in some instrumental exams, therefore ensuring a certain standard of aural skills for every young musician. In Germany, such professional institutionalized systems for the acquisition of theoretical and aural skills are still missing, and the training literature, such as that for ear training, is multitudinous and highly subjective (Estrada Rodriguez, 2008).

Despite this lack of a well-organized curriculum and training, theoretical and aural skills are stringently tested before admission to a university. As far as we know, there has not been any

research concerning the benefits of skills acquired in music theory and ear training for ambitious adolescent musicians and future university of music students. However, the predictability of academic accomplishments and the necessity of theoretical skills in music have been the topic of research.

Harrison (1990) examined the predictability of music theory grades in an American college and found slightly different predictors depending on whether the first or the second year's grade was to be predicted. In both samples the Scholastic Assessment Test (SAT) score in mathematics and the high school grade point average were positive predictors. In the first year's sample, the SAT verbal score and the fact that a person had learned the piano were stable predictors for a good performance in music theory; however, in the second year's sample there was a negative influence for those students whose main instrument was neither the piano nor voice, amongst other factors. Harrison did not discuss this negative predictor of non-pianists and non-singers, although this significantly disadvantageous effect means that pianists and singers possess skills beneficial to the study of music theory—and other instrumentalists do not.

A more recent survey (Papageorgi et al., 2010) on the perception of expertise for music students and musicians investigated the general importance of musical skills, the rating of one's own skills, components of expert performance, and one's own level of expertise. Several items addressed skills in music theory and ear training (i.e., the importance of "acute ear/detailed learning" or "Expert performers are much more competent in reading musical notation") but in the factor analyses these items did not establish clear factors for theoretical, analytical or auditory skills, nor did the authors examine the properties of the items and how well they measured the sought latent variables. Similar findings are reported in other studies concerning the components of musical skills (for a summary, see Hallam, 2006).

Such findings are in contradiction to the curricular plans for most music programs. This could be due to the nature of the research instruments and the samples: Participants who have never learned about music theory will rarely mention it in a qualitative study or rate it high in a survey. Professional musicians are possibly more cognizant of the quality of their performance than of the importance of a musical ear or reading music (Hallam, 2006; Hallam & Prince, 2003). In addition, these skills are not observable concepts of musicianship or emotional involvement; their influence on musical performance is of a more mediate kind.

Musical entrance exams: A challenge to psychometrics

At German universities of music, skills in music theory and ear training are requirements for all candidates. The career aspiration of these candidates wishing to study music is to become a professional instrumentalist, an instrument teacher or a music teacher at a public school. These applicants have to succeed in an entrance exam that always includes an audition in front of a jury as well as a written exam in music theory and ear training. Its purpose is to examine the exceptional skill of the student and to choose those candidates who are the best. To master this exam, it is necessary to have already achieved a certain degree of expertise in the relevant subject areas.

In a previous comparison of such entrance exams (Wolf, Kopiez & Platz, 2012), it has been shown that different universities of music use different test items and do not determine in advance a number of objective criteria in terms of test theory to determine the right solution. This finding is especially worrying for three reasons.

First, the music theory community has not agreed on assessment criteria for musical "essay-like" test items, such as melodic dictation and harmonic or counterpoint work (Gillespie, 2001). Neither the student nor even a colleague can fully reconstruct how the grade has been attained. Such objective assessment criteria are not even locally defined for the exams, although they are

a necessary prerequisite for psychometric assessments (Hornke, 2005; ISO 10667, 2011). Second, two out of the three entrance exam regulations we thoroughly analyzed contain pre-defined guidelines for compensation for poor grades in the entrance exams. An immediate consequence of such a measure is the admission of students who play music brilliantly but possibly lack basic knowledge in notation and aural skills. The overall median grades (use of median due to non-normal distributions) of the music theory entrance exam at one German university of music in three consecutive years were 3, 3 and 5 out of 15 points. However, this underachievement is not an ultimate result, as the grading of the exam is not fully understandable. Third, a test item comparison has shown that each of the three universities of music tests differently. Due to high rejection rates, most applicants apply to several universities of music and therefore have to prepare for different exams. While the aim of the Bologna process was a standardization of all study programs within Europe, not even universities of music within one federal state in Germany use the same items on their music theory entrance exam. This diversity is also in contradiction to the musical practice of competitions and orchestral auditions in which all musicians are required to play the same predetermined pieces.

Wolf et al. (2012) have found a general need for improvement in music theory and ear training entrance exams. However, since musical experts design and grade the exams using their knowledge and experience in this area, they might still find indicators for the student's performance that are measurable in the current test design.

Lehmann (in press) has conducted a study using the archival data of a university of music to assess the predictability of students' grades. While the correlation between entrance exam and final grade for the main instrument showed a medium effect size ($r = .31$), this correlation differed notably between instruments: For string players, the correlation was practically non-existent ($r = .05$) while for pianists the final grade was predictable to a certain degree (correlation $r = .64$). The correlations between the entrance exam and the final grade for music theory and aural training were notably higher ($r = .45$ and $r = .69$, respectively), but as this study is based on data from only one university of music and a degree (i.e., the "diploma"), which in the meantime has been supplanted with the bachelor's and master's degree, the results require further replication.

Rationale of the study

In this study, we conducted the first long-term study regarding the professional development of music students. By means of a regression analysis, we compared the grades received by candidates on the entrance exam and the final grades upon completion of the program for prospective music teachers. We pursued their development in music theory and their main instrument in the course of up to three years by analysing their grades at an interval of one year each. Additionally, with this data we could trace the students' path of development in studying the main instrument and music theory and the correlations in between. Our principal prediction was the following: Institutions assume that entrance exam grades demonstrate that a candidate has the exceptional artistic proficiency needed to become a musician. We therefore hypothesized that the music theory/ear training and main instrument grades from the entrance exam would predict the subjects' grades in the bachelor's program for music education.

Method

Participants

We analyzed the grades of $N = 63$ former students pursuing a Bachelor of Arts degree in music education at a German university of music ("Musikhochschule"). The students had

started their Bachelor of Arts degree program in 2006 or 2007. Within this program, students are prepared to pursue further studies in scholastic music education and become music teachers in schools, but they are also well prepared for a master's degree in music education or musicology. These two cohorts of students form the biggest group of students in one study program and show a rather homogenous background in culture, language, and concept of music. This sample was chosen following implementation of the Bologna process, which meant that the subjects' standard period of study was four years (scheduled graduation in 2010/2011). We could only include the complete data sets of students who, until this point, had pursued their programs according to the respective courses of study. The use of anonymous data was approved by the administration of the university of music; therefore, it was not possible for us to have information about the age or gender of the students.

Research design

The grades from this longitudinal study were allotted at the entrance exam, which included examinations in music theory and auditions on their main instrument, and during the first three years of study. We converted these grades from the German grading system (1: very good, 2: good, 3: satisfactory, 4: sufficient, 5 and 6: insufficient, with possible decimal places of .3 and .7) to a range of continuous scores from 15 points (highest) to 0 points (lowest). Since we could not assume that the grades were interval scaled, we tested only non-parametrically and reported medians and median absolute deviations (Ruppert, 2011). We also had to keep in mind that students study different instruments with various teachers and that different music theory professors have their students write dissimilar exams.

The predictor variables comprised four grades from the entrance exam: In music theory, two separate grades from the written exam (duration: 1 hour) and from the oral exam (duration: approximately 10 minutes) were included. For the main instrument, we included two grades: The applicants received the first grade for playing an audition that they had prepared in advance and a second grade for spontaneous *prima vista* playing of unknown literature (overall duration: approximately 10–15 minutes). We did not include the grades given at the entrance exam for the second and third instrument auditions because we were only interested in a prediction model concerning the main instrument and music theory.

The typical entrance exam for the bachelor's degree program in music education goes as follows: The applicants arrive in the morning and first take part in the written entrance exam in music theory and ear training. At some point during the day, they meet with the examining committee and complete the rest of the exam. The examining committee consists of university professors and lecturers for the examined instruments and music theory/ear training, as well as a committee chair. Unlike their colleagues who are present at the entrance exams for music performance studies or solo classes, the professors who are present at the entrance examinations to the Bachelor of Music education program are most often not the prospective professors the students will take classes with.

In our study, the criterion variables to be predicted were the grades obtained during the study program for both music theory and the main instrument. All partial grades were module grades after two semesters (or one year) of studies in music theory and at the main instrument. All three successive modules in music theory were included (spanning a learning curve of three years), as well as both the grades for the main instrument of the first two modules (spanning a learning curve of two years). Any grades for the main instrument starting in the third year were excluded because these grades were produced within different settings, i.e., within

different modules which resulted in a different number of credit points and therefore could no longer be compared.

We then calculated the mean grades for each skill using the same proportions as for the overall study grade in the bachelor's degree. In music theory, each grade was weighted with five credit points, so the weighted mean grade in music theory was:

$$MT_{agg} = \frac{5 \cdot \text{Music theory 1} + 5 \cdot \text{Music theory 2} + 5 \cdot \text{Music theory 3}}{15} \quad (\text{Eq. 1})$$

For the main instrument the aggregated weighted mean grade was:

$$MI_{agg} = \frac{8 \cdot \text{Main instrument 1} + 5 \cdot \text{Main instrument 2}}{13} \quad (\text{Eq. 2})$$

These aggregated grades for both disciplines were to be predicted by the entrance exam grades and are denoted by *agg*.

Results

Descriptive analysis

First, we tested the data for normal distribution with the Shapiro-Wilk test, and none of the variables were normally distributed (music theory entrance exam, both written and oral: $p < .1$ and all the other grades: $p < .001$). Following this finding and because of the fact that grades are ordinal-scaled, we then narrowed all further analyses to non-parametric testing.

The most striking finding of this data was the high median of most variables (Med. > 9 on a 15-point scale) for all grades in music theory and the main instrument except the written music theory entrance exam (see Table S1 in the supplemental online section). Figure 1 shows that most students received either good or very good grades once they started studying. The data for the main instrument in the third year is missing because of the incomparable modules; thus, the grades amongst students can no longer be compared. Figure 2 displays the distributions of the written music theory entrance exam and the prepared audition for the main instrument; the respective aggregated study grades in the right column show a very narrow distribution of the grades.

A large overall improvement in the grades was notable, especially in music theory: In the written entrance exam the median grade was at 5.4 points and therefore very low; in the first year it was at 12 points, while the lowest received grade was as high as 9 points.

Table 1 displays all the correlations of the analysed grades in music theory and the main instrument. Because of the restrictions of ordinal-scaled and not normally distributed data, Kendall's τ was calculated.

The correlation coefficients range from $\tau = -.02$ (only negative coefficient) via $\tau = .04$ (smallest positive) to $\tau = .80$ (highest positive) showing only positive substantial correlations in the data. We expected high correlations for the grades within each respective domain, for instance, within music theory and within instrumental performance, but not between them. However, both aggregated grades, which gave an overall representation of the student's skills in the

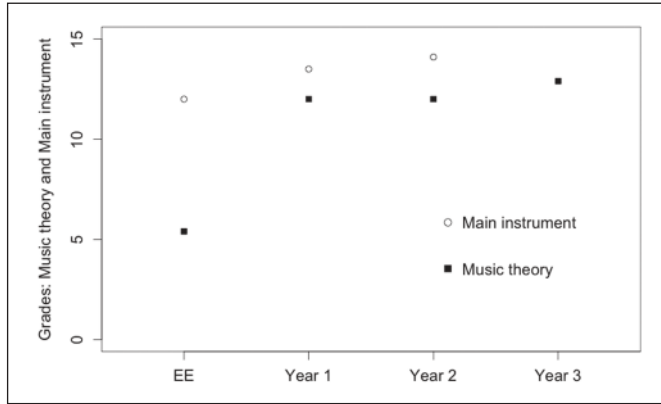


Figure 1. Development of grades in music theory and main instrument exams (median).

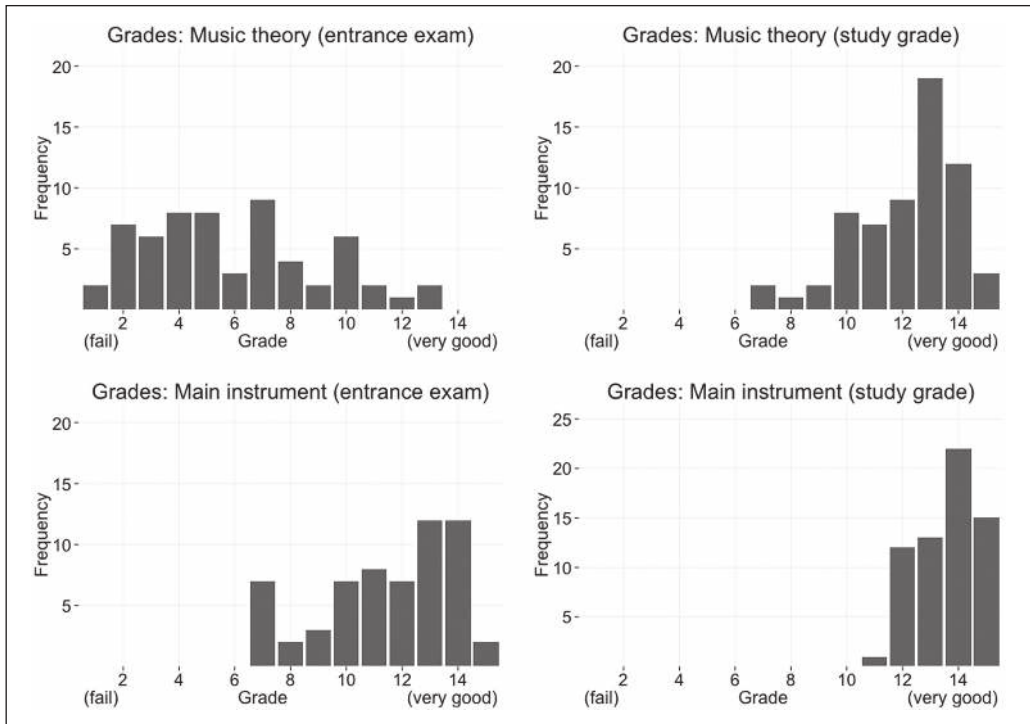


Figure 2. Distribution of the entrance exam and study grades for both music theory and the main instrument.

relevant domain, correlated equally well with both entrance exam grades, irrespective of the subject area (MI agg and MI EE: $\tau = .18$ vs. MI agg and MT EE: $\tau = .19$; MT agg and MT EE: $\tau = .24$ vs. MT agg and MI EE: $\tau = .24$).

Nonetheless, most of the variables displayed a high median and were highly negatively skewed within $G_1 = .415$ (only positive, MT EE) and $G_1 = -1.25$ (mean absolute skewness:

Table 1. Correlation matrix of all analysed grades separated by domains of music theory and main instrument skills, $N = 63$.

	MT EE	MT EE oral	MT 1	MT 2	MT 3	MT agg	MI EE	MI EE prima vista	MI 1	MI 2
MT EE oral	.41*									
MT 1	.23*	.08								
MT 2	.19*	.17	.38*							
MT 3	.23*	.35*	.31*	.39*						
MT agg	.24*	.24*	.61*	.68*	.65*					
MI EE	.19*	.32*	.10	.13	.25*	.24*				
MI EE prima vista	.20*	.27*	.06	.09	.24*	.19*	.39*			
MI 1	.16	.03	.05	-.02	.09	.05	.17	.09		
MI 2	.24*	.06	.12	.14	.14	.17	.18	.13	.33*	
MI agg	.19*	.05	.08	.04	.13	.11	.18	.09	.80*	.57*

Note. Kendall's τ with $\tau \geq .19$ in **bold** ($p < .05$). Abbreviations: MT = Music theory, MI = Main instrument, EE = entrance exam, agg = aggregated final grade.

Table 2. Logistic regression analysis of the predictive quality of the entrance exam grade in music theory (written exam) for the final music theory grade.

Variable	Beta	95% CI	p
Constant	-0.965	0.997	.107
Music theory EE	0.175	0.154	.057
Likelihood Ratio test		$\chi^2 = 3.92$, d.f. = 1, $p = .048$	
Nagelkerke Pseudo R^2		$R^2 = .08$	

$|m_{G1}| = .60$; mean skewness of neg. values: $m_{neg\ G1} = -.619$). This indicated a considerable lack of variance due to ceiling effects in the data. By splitting the sample into high and low achievers in music theory and instrumental performance, respectively, this lack of variance was partly counterbalanced in the next step of the statistical analysis.

Predictive quality of the entrance exam for the final grade in music theory

The sample was divided into two even groups by a median split with regard to the aggregated study grade in music theory. High performers received more than 12.6 out of 15 grade points ($n_H = 31$), and low performers received less than 12.6 grade points ($n_L = 29$). We assumed that this allocation would be well predicted by the entrance exam grade in the same area of skill and as all further exams in music theory were written exams, we only included the grade from the written entrance exam in music theory. By means of a logistic regression analysis, the predictive quality of the entrance exam grade concerning the final (aggregated) grade of the students was to be quantified. It was expected that due to the effort needed for the design and grading of the exams, the entrance exam would effectively explain the allocation to the high- or low-performing group.

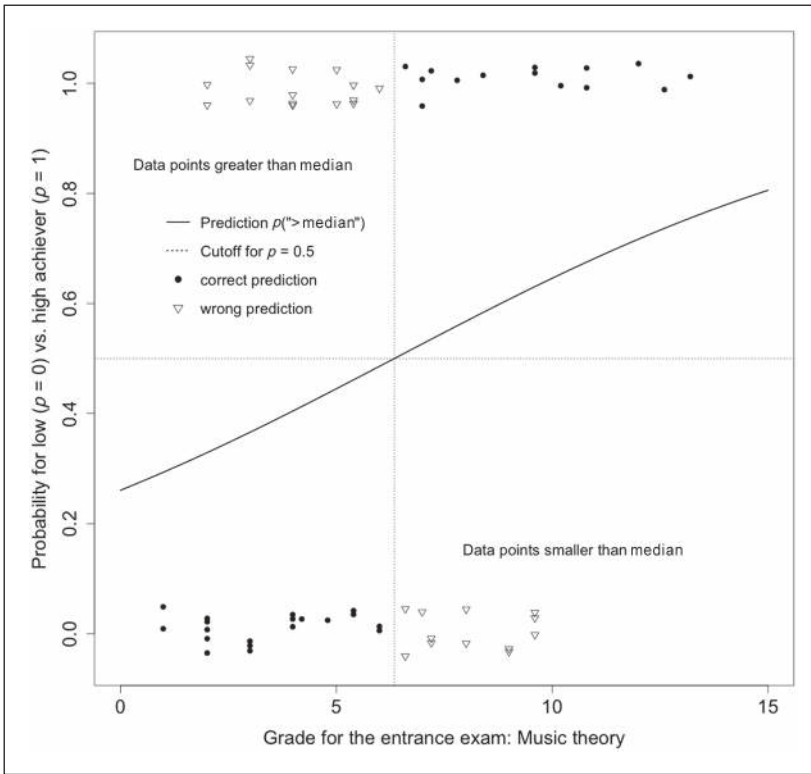


Figure 3. Logistic regression analysis for the development of music theory grades from entrance to final study exam. Allocation of the music theory data to the high- and low-achievement group in the top and the bottom half with data points greater and smaller than median, respectively. All data points on the right of the inflection point $[6, 0.5]$ would have been predicted to be high achievers and those on the left to be low achievers.

This logistic regression (Table 2) produced a prediction model which was significantly different from a null model, so the allocation to the high- or low-performing group was—to a certain degree—predictable by the entrance exam grade. However, it is noteworthy that the shared variance was as low as $R^2 = 8\%$ (Nagelkerke R^2). It became obvious that this model, although significant, provided another indicator for a ceiling effect in the data, general low quality of data, or a problem concerning the prevailing grading system. Figure 3 displays the pattern of both the data points and the prediction of the logistic regression.

Each data point in Figure 3 represents one student, while all data points illustrated by a dot represent a correct prediction and all data points illustrated by a triangle represent a wrong prediction. The figure is divided into four areas: Both top areas include all students predicted to be high achievers; the bottom areas include the students predicted to be low achievers. The students represented by a triangle in the top left area were wrongly predicted to be low achievers by the entrance exam, and they were allocated to the high-performing group by means of the median split. The students represented by a triangle in the bottom right area were wrongly predicted to be high achievers by the entrance exam, and they were allocated to the low-performing group by means of the median split. For both of these groups, each of which contained many students, the entrance exam in music theory proved not to be an ideal assessment of their potential as university students.

Table 3. Logistic regression analysis of the prediction quality of the entrance exam grade (prepared audition) for the main instrument final grade.

Variable	Beta	95% CI	P
Constant	-1.042	2.166	.422
Main instrument e. e.	0.085	0.184	.442
Likelihood Ratio test		$\chi^2 = 0.60$, d.f. = 1, $p = .439$	
Nagelkerke Pseudo R^2		$R^2 = .01$	

Predictive quality of the entrance exam grade for the final grade for the main instrument

To analyze the predictive quality of the entrance exam in the main instrument for the final grade, the sample was divided into two groups by a median split with regard to the aggregated grade for the main instrument. High performers received more than 13.75 grade points ($n_h = 32$), and low performers received less than 13.75 grade points ($n_l = 31$). Here we calculated the regression analysis with the grade for the prepared audition during the entrance exam as all later grades were received for playing prepared pieces.

This logistic regression analysis (Table 3) produced a prediction model that was not significantly different from a null model, so the entrance exam grade did not give any further information about the later achievements in instrumental performance. One possible explanation for this null result is, again, a lack of variance: It is doubtful that a student receiving a very good grade (~13 points) could be considered a “low achiever.” This analysis shows that there was a definite ceiling effect concerning the grades for playing the main instrument (Figure 4). Consequently, the amount of explained variance is negligible ($R^2 = 1\%$).

Specific effects of the main instrument on music theory grades

So far, the results had suggested little success in predicting the study grades with results from the entrance exam. A difference concerning the main instrument was considered: People who play an instrument such as the piano, guitar, organ or accordion have to understand music-related concepts such as chords, harmonies and cadences when practicing. They not only play a melody; they often also play the harmonic framework. Nonetheless, a Wilcoxon rank sum test did not reveal a difference between instrument groups in general performance for either the entrance exam ($W = 440.5$, $p = .825$) or the final grade ($W = 535.5$, $p = .342$) in music theory.

While there was no absolute difference in the quality of grades for students who play different kinds of instruments, a difference in predictability for the final music theory grades was prevalent. For students who play a melodic instrument, such as violin, flute or trumpet, there was a significant correlation between the entrance exam in music theory and the final grade ($\tau = .356$, $z = 2.988$, $p = .003$). For musicians playing a harmonic instrument, there was no significant correlation ($\tau = .070$, $z = 0.453$, $p = .651$).

Specific effects of the main instrument on main instrument grades

This same split of the sample into harmonic and melodic instrumentalists was also applied to the final grades received for playing the main instrument. There was no difference in grading for the main instrument in the entrance exam as measured by the Wilcoxon rank sum test ($W = 480.5$, $p = .405$) nor in the final grade ($W = 454.0$, $p = .848$).

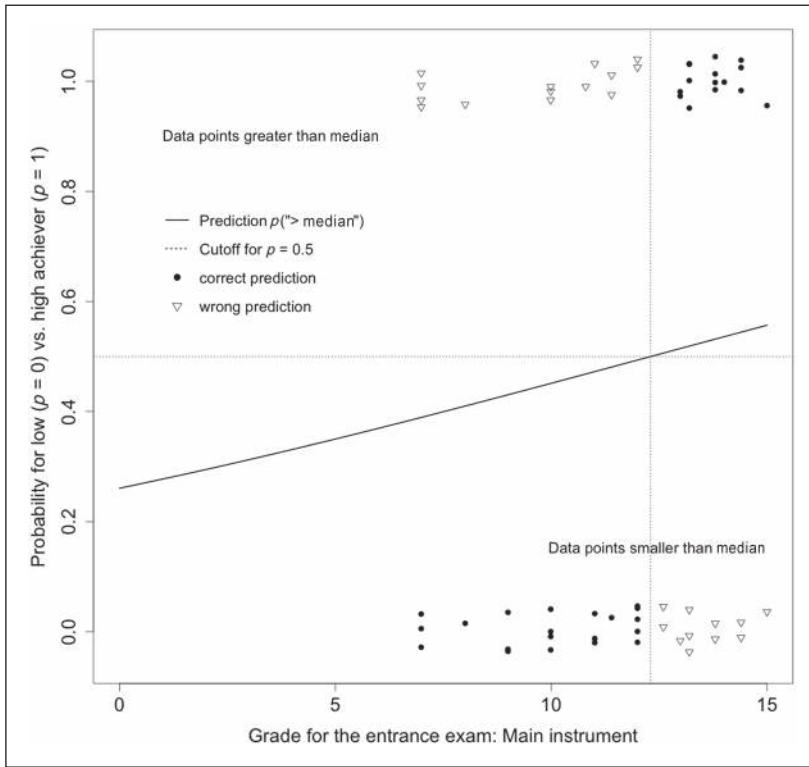


Figure 4. Logistic regression analysis for the development of main instrument grades from entrance to final audition. For further information see Figure 3. The relevant inflection point here is [12, 0.5].

The predictability within each instrument group was again determined for the main instrument. In terms of the main instrument, this entrance exam and the final grade were now significantly correlated for harmonic instrumentalists ($\tau = .294, z = 1.90, p = .058$) while there was no substantive correlation for melodic instrumentalists ($\tau = .138, z = 1.15, p = .251$), revealing a discrepancy between both investigated subject areas. As these findings were not based on a priori hypotheses, these latest results could as well be meaningless artifacts in our sample. However, in future studies such instrument-specific effects should be investigated.

Discussion

This study has shown for the first time that musical entrance exams can lack prognostic validity. This result is based on the longitudinal data of 63 students who we followed for three years. Although these data came from one university of music, they are probably common to universities of music in Germany: The conference of the German directors of universities of music has decided on a number of quality standards, including a stipulation for the entrance exam’s structure: An entrance exam ordinarily entails both an artistic and a theoretical exam to ensure a certain level of skill in both domains (Rektorenkonferenz der deutschen Musikhochschulen in der HRK, 2009), such as is the case for our study. Despite this modest attempt at standardization, each of the 24 German universities of music produces its own entrance exam, including

the instrumental audition as well as music theory and ear training. This is even more peculiar considering that students aim for internationally standardized degrees. Nonetheless, the beginning of this journey toward a degree is an idiosyncratic entrance exam, as there have not been any attempts to agree on well-defined skills or even to standardize this procedure for all of Germany. This variety is not only manifest in entrance exams, but also in music theory and ear training pedagogy, as has been shown by a recent survey amongst most German universities of music (Kühn, 2010; see also Estrada Rodriguez, 2008 and Menke, 2010).

Another reason for the lack of prognostic validity of entrance exams is the subsequent grading during the study program, as students are graded outstandingly high. Such a ceiling effect has only recently been confirmed for German universities in general (German Council of Science and Humanities, 2012) and for musical study programs in particular (Federal Office of Statistics, 2012; Lehmann, in press).

Concerning the music theory and ear training grades, another set of explanations remains plausible: First, it is possible that the students are taught music theory and ear training exceptionally well during their first year of study resulting in the extraordinary performance leap observed (median 5.4 to 12 out of 15). Second, the entrance exam could serve a diagnostic purpose in ascertaining the knowledge of the respective cohort. If this were the case, the curriculum would then be chosen according to what is achievable for the students. This might result in a levelling of the knowledge of most students so that students with poorer grades make up a lot of ground and eventually match the more advanced students. Third, the entrance exam could be designed in such a difficult way that most students fail, so that other measures become important in the process of admission. This “strategic” grading would lead to a devaluation of the music theory exam and the student’s being accepted on the basis of his/her instrumental grade. Since a poor music theory grade can, in this sample, be compensated for by a high instrumental grade, this is a likely explanation. We cannot draw any conclusions about these students who did not pass the entrance exam or about their hypothetical success or failure in studying music. With the current set of data these partly conflicting explanations cannot be answered.

It has been shown by McPherson et al. (1997, see also McPherson, 1995) that musical performance is a complex model of separate skills (sight-reading, performing rehearsed music, playing from memory, playing by ear, and improvising), which are highly positively associated with each other and depend on a number of musical thought processes. Such musical thought processes again rely on a functional match of sound, play, and notation (McPherson & Gabrielsson, 2002). Notably, these three skills have also been adapted by music theory pedagogues to enable musicians to “think in music,” a skill they manifest when they understand what they hear, audiate, and understand notation they read (Karpinski, 2000). Yet, as shown by Brodsky et al. (2003) this skill requires extensive training and is no mere by-product of instrumental expertise.

The study of one subskill of musical performance, the sight-reading of new music, has provided a number of skill components predicting the accomplishments in sight-reading. Amongst these components are practice-related and -unrelated skills; pursuant to a pedagogical approach to make every student reach his or her full potential, the practice-related skills are more applicable for everyday music-making and the education of ambitious musicians. These practice-related skills consist of two important components: sight-reading expertise by the age of 15 and the ability of inner hearing. The first is adoptable to the weekly instrumental lessons; the second can be assumed to improve with regular aural skills training. Due to the impact of inner hearing, sight-reading is a skill that improves from a well-trained combination of reading, hearing, and playing skills (Kopiez & Lee, 2008).

The present study examined the prognostic validity of a musical entrance exam by investigating its impact on the respective grades received in the study program. While Lehmann (in press) has provided evidence for differences in predictability for different instruments, we could not replicate his finding. Such effects might be university-specific or more specific for the highly specialized music students in comparison to the more broadly qualified music education students.

Yet, we could draw some cautious conclusions for different instrument groups: While the final grade for the main instrument of harmonic instrumentalists could be predicted to a certain degree, the music theory grade could not. And while the grade for the main instrument of melodic instrumentalists could not be predicted, the music theory grade could. Therefore, only one of the analyzed entrance exams was useful for the prediction of the students' success in their program of study, and it was a different one for different instrument groups. The discrepancies in the prediction for music theory might be explained in this way: Pianists, guitarists and other musicians dealing with harmonies and harmonic progressions on a daily basis might have acquired sufficient skills that do not need to be comprehensively improved. Melodic instrumentalists, on the other hand, have not yet achieved enough understanding for music theory and learn many concepts for the first time in their course of studies. Their learning curve might be easier, so systematic improvements are more likely than for harmonic instrumentalists.

Nonetheless, we did replicate the general grade inflation from Lehmann (in press) and a medium correlation for the main instrument grades (entrance exam and final grade) for the subgroup of harmonic instrumentalists. These discrepancies can be ascribed to the differences between the investigated universities of music and to the degree system: The grades from our sample originate from those students who were already in the Bachelor and Master degree programs, while the data from Lehmann were produced within the diploma degree system. This difference is mostly manifested in the difference of one set of final diploma examinations compared with the successive examinations of the bachelor system. In contrast to the study by Harrison (1990), who also examined the predictability of grades in degree programs, Lehmann's and our studies only utilized grades within one domain (music theory/ear training or instrumental skills). It has not yet been theoretically established how transfer effects from mathematics or a general intelligence factor, for example, can explain a student's achievements in music theory, yet the knowledge in one domain at several stages of development should show some cohesion.

Limitations of the study

The post-hoc test power of this study with $N = 63$ participants is moderate. For the established "5–20-rule," with $\alpha = 0.05$ and $1 - \beta = 0.8$, effect sizes of Cohen's $d \geq 0.32$ should have been detected in this setting (Faul, Erdfelder, Buchner, & Lang, 2009). However, there were several reasons to not expand the sample. First, students from other study programs would not have had the same class structure, just as students from the related program before the conversion from the previous degree "state examination" (Staatsexamen) to the Bachelor of Arts degree would not. The integration of students from other universities would have led to the same problems. Despite this detriment, the occurring substantial misallocation of predicted "high" and "low" performers is not likely to disappear in a larger sample of students simply due to the law of large numbers and the reduction of random error (here: random misallocation). We also assume that similar results occur at other (German) universities of music because examinations are, to our knowledge and as Kühn (2010) writes, designed on the basis of examiners' subjective experience at each university. While this educational sovereignty is the foundation of

a diverse and multifaceted culture, prospective students would still benefit from knowing the standards of their disciplines.

Given that similar results can be replicated at other universities of music, it will be useful to consider a change of learning and testing procedures prior to university admission. Universities of music would have to define the requirements for future students, enabling young people to learn these skills in a professional manner before applying to a university. Despite the federalist organization of public education in Germany, the final high school exams are already in the process of being standardized. A comparable successive standardization of tests is also applicable to music theory and ear training entrance exams, as has already been proposed (Estrada Rodriguez, 2008). This is even more reasonable considering that students will receive the same degree at every university of music after successful examinations and play standardized repertoire when applying for jobs at orchestras or opera houses.

Comparable to the expenses associated with studying medicine, each university admission to degree programs in drama or music is very costly due to the private lessons and small group classes. It is therefore necessary to identify the best students from all applicants efficiently and with selection procedures that best predict success in the degree program.

Suggestions for future studies

Notably, the testing procedure currently in use is far from optimal for any of the parties involved: Prospective students need to learn different subskills for different universities of music and have no professional learning environment to guide them. Music theorists design a test which is a poor predictor for study success and, due to its high difficulty for the students, is often partly ignored (“compensated for”) by the program coordinator when admitting students for a program.

In concurrence with the educational trend to agree on fundamental competencies and standards in all study subjects and assessments in general (Hornke, 2005; ISO 10667, 2011), we suggest initiating a similar process for the subjects of music theory and ear training as well as musical skills in general. An assessment instrument of expert music performance will enable objective and valid identification of the students with the best prerequisites for study and reveal to instructors skill differences in the saturation range of professional musicians.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Associated Board of the Royal Schools of Music. (2012). *Music theory exams and Syllabus*. Retrieved from <http://www.abrsm.org/en/our-exams/music-theory-exams/>
- Brodsky, W., Henik, A., Rubinstein, B.-S., & Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. *Perception & Psychophysics*, 65(4), 602–612.
- College Board. (2012). *Advanced placement in music theory, Course description*. Retrieved from <https://apstudent.collegeboard.org/apcourse/ap-music-theory/course-details>
- Estrada Rodriguez, L. A. (2008). *Didaktik und Curriculumentwicklung in der Gehörbildung. Eine vergleichende Untersuchung an deutschen Lehrbüchern zur Gehörbildung aus der Zeit 1889 bis 1985 [Didactics and curriculum development in aural training. A comparative study of German textbooks for ear training between 1889 and 1985]*. Hannover: Institute for Research in Music Education.
- Fallows, D. (2011). Theory. In *Oxford Companion to Music*. Retrieved from <http://www.oxfordmusiconline.com/>

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149–1160.
- Federal Office of Statistics. (2012). *Bildung und Kultur. Prüfungen an Hochschulen* [Education and culture. Examinations at universities]. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PruefungenHochschulen2110420117004.pdf?__blob=publicationFile
- Fry, C., & Spencer, P. (2011). Ear-training. In *Oxford Companion to Music*. Retrieved from <http://www.oxfordmusiconline.com/>
- German Council of Science and Humanities. (2012). *Prüfungen an Hochschulen im Prüfungsjahr 2010* [Grades at colleges and universities in the 2010 academic year]. Retrieved from <http://www.wissenschaftsrat.de/download/archiv/2627-12.pdf>
- Gillespie, J. (2001). Melodic dictation scoring methods: An exploratory study. *Journal of Music Theory Pedagogy*, 15, 51–68.
- Hallam, S. (1998). *Instrumental teaching: A practical guide to better teaching and learning*. Oxford: Heinemann.
- Hallam, S. (2006). Musicality. In G. E. McPherson (Ed.), *The child as musician. A handbook of musical development* (pp. 93–110). New York: Oxford University Press.
- Hallam, S., & Prince, V. (2003). Conceptions of musical ability. *Research Studies in music Education*, 20(2), 2–22.
- Harrison, C. S. (1990). Predicting music theory grades: The relative efficiency of academic ability, music experience and musical aptitude. *Journal of research in music education*, 38(2), 124–137.
- Hornke, L. F. (2005). Die englische Fassung der DIN 33430 [The English version of the DIN 33430]. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K.D. Kubinger, F. Lang, H. Moosbrugger, et al. (Eds.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*. Lengerich: Pabst. Retrieved from <http://www.bdp-verband.org/bdp/politik/clips/din33430en.pdf>
- ISO 10667. (2011). *Assessment service delivery—Procedures and methods to assess people in work and organizational settings*. Berlin: Beuth.
- Karpinski, G. S. (2000). *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians*. Oxford: Oxford University Press.
- KoMus—Competence model for music. (2009). *Project description*. Retrieved from <http://www.musik.uni-bremen.de/de/forschung/musikpaedagogische-forschungsprojekte/komus-kompetenzmodell-im-fach-musik/project-komus-english-version.html>
- Kopiez, R., & Lee, J. I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, 8(1), 97–120.
- Kopiez, R., & Lee, J. I. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, 10(1), 41–62.
- Kühn, C. (2010). Musiktheorie lehren. Zu einer Umfrage an den deutschen Musikhochschulen [Teaching music theory. A survey conducted at German universities of music]. *Zeitschrift der Gesellschaft für Musiktheorie*, 7(2). Retrieved from <http://www.gmth.de/zeitschrift/artikel/505.aspx>
- Lehmann, A. C. (in press). Predicting final grades for principal instrument, aural skills, and music theory in a college music program from admission assessments. *Journal of Research in Music Education*.
- Menke, J. (2010). Brauchen wir einen Kanon in der Musiktheorie? [Do we need a music theory canon?] *Zeitschrift der Gesellschaft für Musiktheorie*, 7(2). Retrieved from <http://www.gmth.de/zeitschrift/artikel/507.aspx>
- McPherson, G. E. (1995). Five aspects of musical performance and their correlates. *Bulletin of the council for research in music education*, 127, 115–121.
- McPherson, G. E., Bailey, M., & Sinclair, K. E. (1997). Path analysis of a theoretical model to describe the relationship among five types of musical performance. *Journal of Research in Music Education*, 45(1), 103–129.
- McPherson, G. E., & Gabriellson, A. (2002). From sound to sign. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance: Creative strategies for teaching and learning* (pp. 99–115). Oxford: Oxford University Press.

- National Assessment of Educational Progress. (2008). *NAEP Arts assessment*. Retrieved from <http://nces.ed.gov/nationsreportcard/arts/>
- Papageorgi, I., Creech, A., Haddon, E., Morton, F., F. Lang, H. Moosbrugger, et al. (2010). Perceptions and prediction of expertise in advanced musical learners. *Psychology of Music*, 38(1), 31–66.
- Rektorenkonferenz der deutschen Musikhochschulen in der HRK. (2009). *Leitbild der deutschen Musikhochschulen [General mission of German universities of music]*. Retrieved from <http://www.die-deutschen-musikhochschulen.de/ueber-uns/leitbild/>
- Ruppert, D. (2011). *Statistics and data analysis for financial engineering*. New York: Springer.
- Wolf, A., Kopiez, R., & Platz, F. (2012). Der Status Quo der musiktheoretischen Zulassungsprüfung an Musikhochschulen: Eine testtheoretische Analyse [The status quo of collegiate entrance exams in music theory analysed from the perspective of test theory]. *Beiträge empirischer Musikpädagogik*, 3(2). Retrieved from <http://www.b-em.info/index.php?journal=ojs&page=article&op=viewFile&path%5B%5D=77&path%5B%5D=220>